UNIVERSIDADE DE LISBOA

Faculdade de Psicologia

Beyond discrete biases: Human judgment as a continuous interactive process

Joana Filipa Figueiredo Ribeiro dos Reis

Orientadores: Prof. Doutor Mário Augusto de Carvalho Boto Ferreira
Prof. Doutor Leonel Garcia-Marques
Prof. Doutor Florian Kutzner

Tese especialmente elaborada para a obtenção do grau de Doutor em Psicologia,

especialidade de Cognição Social

2020

UNIVERSIDADE DE LISBOA

Faculdade de Psicologia

Beyond discrete biases: Human judgment as a continuous interactive process

Joana Filipa Figueiredo Ribeiro dos Reis

Orientadores: Prof. Doutor Mário Augusto de Carvalho Boto Ferreira
Prof. Doutor Leonel Garcia-Marques
Prof. Doutor Florian Kutzner

Tese especialmente elaborada para a obtenção do grau de Doutor em Psicologia, especialidade de Cognição Social

Júri:

Presidente: Doutora Ana Margarida Vieira da Veiga Simão, Professora Catedrática e Membro do Conselho Científico da Faculdade de Psicologia da Universidade de Lisboa;

Vogais:

- Doutor José Bernardo Bicudo de Azeredo Keating, Professor Associado da Escola de Psicologia da Universidade do Minho;
- Doutora Margarida e Sá de Vaz Garrido, Professora Associada da Escola de Ciências Sociais e Humanas do ISCTE – Instituto Universitário de Lisboa;
- Doutor Mário Augusto de Carvalho Boto Ferreira, Professor Associado da Faculdade de Psicologia da Universidade de Lisboa, orientador;
- Doutor André Otelo Paraíba Mata, Professor Auxiliar da Faculdade de Psicologia da Universidade de Lisboa.

Investigação financiada pela Fundação para a Ciência e a Tecnologia através da bolsa de doutoramento SFRH/BD/100736/2014

2020

The research reported on this thesis was funded by

Fundação para a Ciência e a Tecnologia,

Doctoral Grant SFRH/BD/100736/2014

FCT

Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Acknowledgments

Four years ago I was ten years younger. When I decided to make a PhD on judgment and decision making under uncertainty I was far from guessing how challenging it would be, precisely because I am the worst person to deal with uncertainty (particularly when combined with making judgments and decisions).

Along this journey, I often battled with my options but I was fortunate to have people around me who guided and encouraged me all the way through.

My first badge goes to Prof. Mário Ferreira, my main advisor, for all his guidance and encouragement but also for his enormous patience to deal with my uncertainty avoidance. Thank you Mário for promoting my personal and professional development by encouraging me to explore my own ideas and by subtly pushing me out of my comfort zone.

The second badge goes to Prof. Leonel Garcia-Marques, my Portuguese co-advisor, for his always valuable feedback - and feedforward - on this work but also for his inspiring role within the Social Cognition and Learning research group.

I would also like to thank Prof. Florian Kutzner, my German co-advisor, for all his support during my stay in Heidelberg. In this regard, I would like to extend my gratitude to Prof. Klaus Fiedler and the researchers in CRISP.

I am also thankful to all my colleagues in the Social Cognition and Learning research group and those with whom I shared the A321 room for making it possible to share ideas, achievements and disappointments. I am particularly thankful to Prof. André Mata for his always sharp comments and suggestions, and for his contagious research enthusiasm; to Cristina Mendonça for her invaluable tips and help in programming some of the present experiments, and for helping to combat my writing procrastination, together with Ana Lapa and Sara Hagá; and to all the students, colleagues and lab managers that helped me with the logistics of data collection.

On a more personal note, I would like to thank the unrelenting support and precious advices of Ana Rita Farias, Cláudia Simão, Patrícia Pascoal and Rita Jerónimo who have already been through the PhD ordeal.

Lastly, I would like to thank my family and friends for their love and unconditional support. Thank you all for waiting for me at the finish line and for always reminding me there is life out there.

Lisbon, 18th October 2019

Resumo

A presente tese tem como objetivo testar o papel dos ambientes de decisão na correção ou manutenção dos enviesamentos de julgamento e decisão decorrentes do uso de heurísticas. Usando como ponto de partida a abordagem funcional das heurísticas de acordo com a qual a qualidade do julgamento é determinada pela qualidade do feedback providenciado, procurou testar-se de que forma as respostas (enviesadas) que as pessoas dão em tarefas clássicas de resposta única ou discreta (e.g., pseudodiagnosticidade, ancoragem, problemas raciocínio que envolvem um conflito entre uma resposta heurística e uma resposta deliberada) evoluem em função do feedback (social ou não social) recebido em ambientes de aprendizagem contínuos (i.e., ambientes de resposta sequencial a sucessões de problemas). Propõe-se que este feedback pode contribuir para corrigir os enviesamentos quando, a) em ambientes de aprendizagem contínuos, é imediato e válido, revelando a inadequação das estratégias usadas (Capítulo 2); ou quando b) em contextos sociais, as pessoas são expostas a respostas inconsistentes com as suas, dadas por outros percebidos como pouco competentes ou enviesados (Capítulos 3 e 4). Pelo contrário, os enviesamentos tendem a persistir quando os ambientes de decisão mantêm as pessoas erradamente convencidas da adequação das suas estratégias. Por exemplo, quando o feedback recebido é incompleto ou enganador (Capítulo 2) ou quando os outros são percebidos como competentes e/ou as suas respostas coincidem com as respostas baseadas em heurísticas que a maioria dos participantes ingénuos dá por defeito (Capítulo 4).

Palavras-chave: heurísticas e enviesamentos; ambientes de decisão kind e wicked; vigilância epistémica; influência social; ambientes de aprendizagem contínuos.

Abstract

This thesis aims to test the role that decision environments might play in the correction or maintenance of judgment and decision biases resulting from the use of heuristics. Building on the functional approach to heuristics, according to which the quality of judgments is determined by the quality of the feedback provided, we sought to test how the (biased) answers that people give in classic one-shot tasks (e.g., pseudodiagnosticity, anchoring, reasoning problems that involve a conflict between an heuristic and a deliberate response) evolve as a result of feedback (social or non-social) received in continuous learning environments (i.e., environments where people answer to several problems that are sequentially presented). It is proposed that feedback can contribute to correcting biases when, a) in continuous learning environments, it is immediate and valid, revealing the inadequacy of the strategies used (Chapter 2); or when (b) in social contexts, people are exposed to responses that are inconsistent with their own, given by others perceived as not competent or biased (Chapters 3 and 4). In contrast, biases tend to persist when decision environments keep people wrongly convinced of the appropriateness of their strategies. For example, when feedback received is incomplete or misleading (Chapter 2) or when others are perceived as competent and / or their answers match the heuristic-based responses that most naive participants already give by default (Chapter 4).

Keywords: heuristics and biases; kind and wicked decision environments; epistemic vigilance: social influence; continuous learning contexts

Resumo alargado

O programa de investigação em heurísticas e enviesamentos (e.g., Tversky & Kahneman, 1974) revolucionou o estudo do raciocínio, julgamento e tomada de decisão ao demonstrar que o julgamento humano se baseia em heurísticas (i.e., formas simplificadas de tomada de decisão) que são qualitativamente diferentes das prescrições dos modelos normativos de decisão racional. A tese inicial deste programa de investigação era que embora eficazes em muitas circunstâncias, estas heurísticas conduzem também a enviesamentos sistemáticos. Estes enviesamentos têm sido objeto de estudo intenso e amplo nas últimas cinco décadas (e.g., Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982).

Curiosamente, apesar de ser unanimemente reconhecida a natureza contínua e social dos ambientes em fazemos julgamentos e tomamos decisões, a maior parte da investigação na linha do programa de investigação em heurísticas e enviesamentos focou-se nas respostas e soluções que as pessoas dão em tarefas inferenciais “one-shot” ou pontuais (i.e., tarefas onde os participantes procedem à resolução discreta ou isolada de um ou mais problemas), negligenciando o modo como estas respostas se alteram e evoluem quando as pessoas têm oportunidade de ajustar progressivamente as suas estratégias de julgamento em função do feedback do ambiente. A presente tese tem como objetivo contribuir para preencher esta lacuna na investigação testando empiricamente o papel que os ambientes de decisão podem ter na correção ou manutenção de erros e enviesamentos de julgamento e decisão.

Teoricamente enquadrada pelas teorias dualistas de julgamento e decisão (para uma revisão, ver Evans, 2007), a investigação aqui apresentada assenta no pressuposto de que o sistema cognitivo humano é suficientemente flexível para permitir que as pessoas baseiem os seus julgamentos em processos mais heurísticos (tipicamente referidos como processos Tipo 1 - T1) ou mais deliberados (normalmente referidos como processos Tipo 2 - T2 - na nomenclatura da abordagem dualista) em função do feedback recebido do ambiente (e.g., Ferreira, Garcia-Marques, Sherman & Sherman, 2006). Nesse sentido, propõe-se que os enviesamentos frequentemente encontrados em tarefas discretas na tradição do programa de heurísticas e enviesamentos (maioritariamente associados a um processamento T1) podem ser corrigidos, em ambientes de decisão que permitam às pessoas reconhecer os seus erros ou a inadequação das suas estratégias de julgamento, levando-as a adotar um modo de processamento deliberado (T2). Pelo contrário, estes enviesamentos deverão persistir em ambientes que mantenham as pessoas (erradamente) convencidas da adequação das suas estratégias.

Usando a abordagem funcional das heurísticas (Hogarth, 1981, 2001) como ponto de partida, os estudos realizados focam-se em dois aspetos dos ambientes de decisão relativamente negligenciados na investigação anterior. Em primeiro lugar, explora-se em que medida o feedback que as pessoas recebem enquanto respondem a problemas lhes permite tomar consciência da inadequação das suas estratégias, levando-as a adotar um modo de processamento mais elaborado ou, pelo contrário, reforça o uso de estratégias intuitivas inadequadas ao mantê-las convencidas da sua eficácia.

Nas experiências incluídas no Capítulo 2 usou-se uma versão do paradigma clássico de pseudodiagnosticidade (e.g., Doherty, Schiavo, Tweney, & Mynatt, 1981) adaptada para ambientes contínuos. Com o objetivo de testar em que medida as pessoas modificam as suas estratégias de seleção de informação em função do feedback recebido do ambiente, foram criados ambientes *kind* e *wicked* (Hogarth (2001). Ambientes kind são ambientes em que o feedback dado aos participantes revela claramente a inadequação das suas estratégias; ambientes wicked são ambientes cujo feedback mantém os participantes erradamente convencidos da qualidade das suas estratégias.

Os resultados destas experiências demonstram que o feedback fornecido em ambientes de aprendizagem contínuos permite reduzir o uso de estratégias intuitivas de seleção de informação ao longo dos ensaios. Enquanto na Experiência 1 esta redução foi maior no ambiente kind do que no wicked, na Experiência 2 a diferença entre os dois ambientes não foi significativa. No entanto, o desempenho dos participantes num problema mais complexo apresentado no final da experiência sugere que o feedback fornecido no ambiente kind melhorou a compreensão profunda da tarefa, enquanto a melhoria de desempenho observada no ambiente wicked mostrou ser mais superficial, não se transferindo para o novo problema.

Por outro lado, tendo como fonte de inspiração a teoria da elaboração do conflito (Mugny, Butera, Sanchez-Mazas, & Perez, 1995) e os mecanismos de vigilância epistémica propostos pela teoria argumentativa do raciocínio (Mercier & Sperber, 2011; Sperber et al., 2010) procura-se testar em que medida a exposição a respostas de outras pessoas - apresentadas em contextos sociais - pode contribuir para desencadear um processamento mais elaborado. Propõe-se que a exposição a respostas de outras pessoas ativa um tipo de processamento mais elaborado quando as pessoas têm razões para desconfiar dessas respostas – por exemplo, quando são dadas por uma fonte pouco competente ou enviesada e divergem das respostas que as pessoas dariam por si próprias. Pelo contrário, as mesmas respostas são adotadas sem grande escrutínio quando: 1) são dadas por outros percecionados como muito competentes; ou 2) coincidem com as respostas intuitivas (baseados em heurísticas) que as pessoas dariam por si próprias, independentemente da competência da fonte.

No Capítulo 3 são apresentadas duas experiências nas quais se usou uma versão modificada do paradigma de ancoragem (Strack & Mussweiler, 1997). Procurou-se testar em que medida a tendência para percecionar os outros como mais enviesados do que o próprio (Pronin, Lin, & Ross, 2002), leva as pessoas a adotar um modo de processamento mais elaborado quando avaliam as estimativas (i.e., âncoras) de outros, por comparação com as mesmas estimativas apresentadas sem uma fonte social identificada.

Nas duas experiências foi pedido aos participantes que respondessem a um conjunto de perguntas de conhecimento geral, depois de considerarem possíveis respostas alegadamente dadas por outros participantes ou apresentadas sem uma fonte identificada. Dependendo da condição experimental, antes de responderem a estas perguntas os participantes eram avisados (ou não) da possibilidade dos julgamentos serem influenciados pelas respostas apresentadas (efeito ancoragem). Estes avisos deveriam levar os participantes a tratar as respostas dos outros participantes como enviesadas, desencadeando assim um modo de processamento mais elaborado, conducente à redução do efeito. Pelo contrário, quando as mesmas respostas eram apresentadas sem fonte ou quando os participantes não recebiam um aviso do efeito, as suas estimativas deveriam manter-se mais ancoradas às respostas apresentadas.

Os resultados destas experiências revelaram uma redução do efeito de ancoragem quando os participantes foram avisados deste enviesamento e as âncoras foram apresentadas como respostas de outros participantes. Em linha com o proposto, a medição dos tempos de resposta na Experiência 1 sugere que esta redução do enviesamento resulta de processos deliberados de ajustamento a partir das âncoras dadas.

Finalmente, as experiências incluídas no Capítulo 4 testam o papel que as respostas de outras pessoas podem ter no desencadeamento de um processamento mais elaborado através da manipulação da competência da fonte destas respostas (Experiências 1 e 2) e da semelhança entre as respostas apresentadas e as respostas que a maioria das pessoas dá neste tipo de tarefas (Experiência 2). Em ambas as experiências, foram usados problemas de raciocínio que apresentam um conflito entre respostas heurísticas, altamente apelativas mas erradas, e respostas deliberadas, com base em regras e válidas. Foi pedido aos participantes que respondessem aos problemas depois de considerarem as respostas de outros (alegados) participantes descritos como muito ou pouco competentes. Na Experiência 1 estas respostas eram sempre heurísticas, correspondentes às que a maioria das pessoas daria. Na Experiência

2 metade das respostas apresentadas eram heurísticas e metade eram deliberadas. De forma a testar o impacto que as respostas dos outros participantes podem ter nos processos de raciocínio e julgamento individuais, no final de ambas as experiências, foi incluído um bloco de problemas apresentado sem as respostas de outros participantes.

Na Experiência 1 os resultados revelaram uma tendência para seguir as respostas (heurísticas) dos outros participantes, independentemente da sua competência. Pelo contrário, na Experiência 2 os participantes seguiram com mais frequência as respostas de outros descritos como muito competentes. Na condição de elevada competência, os participantes tiveram um melhor desempenho quando as respostas do outro eram as deliberada do que as heurísticas. Contudo, na condição de baixa competência, não se encontraram diferenças entre o desempenho dos participantes após as respostas deliberadas e heurísticas.

Os resultados de ambas as experiências sugerem que a tendência para seguir a resposta dos outros prejudicou o desempenho dos participantes, tendo este melhorado quando, no segundo bloco de problemas, os participantes deixaram de ter acesso às respostas de outros. Outros resultados revelaram ainda a existência de um impacto metacognitivo da exposição às respostas de outros. Ao contrário do padrão habitualmente obtido neste tipo de estudos – maior confiança nas respostas certas do que nas erradas (De Neys, Cromheeke, & Osman, 2011; Mata, Ferreira, & Sherman, 2013) - os participantes de ambas as experiências mostraram-se menos confiantes nas suas respostas certas quando estas divergiam das respostas de outros descritos como muito competentes. Quando as respostas divergiam das de outros descritos como pouco competentes, os participantes mostraram-se igualmente confiantes nas suas respostas certas e erradas.

As implicações destes três conjuntos de estudos serão discutidas e enquadradas à luz da investigação recente em heurísticas e enviesamentos.

Table of Contents

Overview

The heuristics and biases research program (hereafter HB; Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974) indelibly changed Judgment and Decision Making (hereafter JDM) research by proposing that judgment under uncertainty is typically controlled by a limited number of heuristics (cognitive shortcuts) rather than by the formal laws of logic and probability. Despite very efficient under most circumstances, these heuristics were said to result in a variety of systematic biases in many others (Einhorn & Hogarth, 1981; Fischhoff, Slovic, & Lichtenstein, 1977; Gilovich, Griffin, & Kahneman, 2002; Kahneman et al., 1982; Keren & Teigen, 2004; Koehler & Harvey, 2008; Nisbett & Ross, 1980; Slovic, Fischhoff, & Lichtenstein, 1977). For instance, people often ignore alternative hypothesis when evaluating data (e.g., Beyth-Marom & Fischhoff, 1983; Doherty, Mynatt, Tweney, & Schiavo, 1979), let their choices be influenced by irrelevant information (e.g., Englich, Mussweiler, & Strack, 2006; Strack & Mussweiler, 1997), and evaluate the quality of arguments based on their prior beliefs instead of following logical rules (e.g., Evans, Barston, & Pollard, 1983; Evans, Newstead, Allen, & Pollard, 1994;; Klauer, Musch, & Naumer, 2000; Markovits & Nantel, 1989; Thompson & Evans, 2012).

Given the potential negative effects that such biases might have on people's decisions, research in this field soon broaden its focus from the earlier study of specific heuristics and the conditions under which their use resulted in biased answers, to the study of heuristic reasoning, and the definition of the boundary conditions under which it produces errors vs. accurate judgments (Kahneman & Tversky, 1982a; Tversky & Kahneman, 1983).

The dual-process theories of reasoning (e.g., Barbey & Sloman, 2007; Epstein, 1994; Evans, 2008; Evans & Over, 2013; Evans & Stanovich, 2013; Kahneman, 2011; Kahneman & Frederick, 2002; Sloman, 1996, 2014; Stanovich, 1999, 2011, 2018) that are currently dominant in the explanation of these judgmental biases evolved from this earlier research.

According to these theories, when people are faced with a reasoning problem similar to those typically used in the HB field, a default heuristic-based answer (associated with Type 1 processing) comes to mind rapidly and with little effort. This answer is later endorsed, corrected or overridden by the intervention of a more effortful type of processing (Type 2 processing). Judgmental biases are typically attributed to monitoring failures of T2 in detecting T1 biased outputs.

Most previous research aimed at better understanding when and how these failures occur have been focused on variables such as the individual characteristics of the decision maker (e.g., cognitive abilities and thinking dispositions, mood; De Neys, 2006; Stanovich, 1999) or the characteristics of the task (e.g., the amount of time allocated to the task, instructions provided; Evans & Curtis-Holmes, 2005; Finucane, Alhakami, Slovic, & Johnson, 2000), and left largely unexplored the role that the decision environments might play in the regulation of the interplay between T1 and T2 processing.

Starting from Hogarth's (1981) seminal idea that judgmental biases found in one-shot tasks may result from processes that are adaptive and functional in continuous and socially enriched contexts, the present thesis aims to empirically test the role that such 'continuous and socially enriched' decision environments might play in the correction (or reinforcement) of the judgmental biases found in the HB tradition.

Specifically, the ideas explored in the current work derived from three distinct theoretical approaches and the correspondent research programs. First, and as aforementioned, this work was inspired by Hogarth's functional approach to judgmental heuristics and his research program (1981, 2001; Hogarth, Lejarraga, & Soyer, 2015; Hogarth & Soyer, 2011; Soyer & Hogarth, 2015). Hogarth and collaborators argued and showed that the quality of people's (intuitive) decision processes is determined, in the long-run, by the quality of feedback provided by their decision environments. Whereas the accurate and timely feedback offered by the so called kind decision environments provides good opportunities for learning, incomplete or misleading feedback found in wicked environments difficults learning (e.g., Hogarth, 2001; the terms kind and wicked environments will be more thoroughly defined and discussed later on, in the Introduction, pp.19-21).

Another source of inspiration was Mugny and colleagues' (Mugny, Butera, Sanchez-Mazas, & Perez, 1995, see also Pérez & Mugny, 1993, 1996) conflict elaboration theory of social influence. According to this theory, the exposure to external information divergent from individual information that people already hold creates a conflict that leads people to question the validity of their divergent information. In face of this conflict, people tend to accept divergent information provided by a high-status source (e.g., a majority, a high competence source) as more valid than their own, therefore adopting it without further consideration. On the contrary, when the same information is provided by a low-status source (e.g., a minority, a low competence source), people are reluctant to accept it without firstly engaging in a deeper processing of the task (Butera, Legrenzi, Mugny, & Perez, 1992; Butera, Mugny, Legrenzi, & Perez, 1996; Legrenzi, Butera, Mugny, & Pérez, 1991; Nemeth, 1986).

The third source of inspiration was Sperber and collaborators' (Mercier & Landemore, 2012; Mercier & Sperber, 2011, 2017; Sperber et al., 2010) proposal that humans have a suite of cognitive mechanisms for epistemic vigilance to ensure that communication remains advantageous despite the risk of being incidentally or intentionally misinformed. In other words, rather than uncritically accepting the information provided by others, people are vigilant towards the trustworthiness of its source (i.e., how competent and/or benevolent it is), and towards its coherence with their prior knowledge or beliefs. This proposal and the research work it ensued, is based on the more general claim (put forward by Sperber and collaborators) that reasoning is for communication and thus many errors and biases identified by the HB research program occur because studies in this domain have been made, for the most part, in non-social settings where communication between peers is largely absent. Although such strong theoretical assertion is certainly debatable (e.g., Argote, Seabright, & Dyer, 1986; Mowen & Gentry, 1980; Plous, 1993; Tindale, 1993), the empirical questions it raises call for more research on how HB work in social environments (e.g., where one's judgments are shared with or communicated to other). To shed some light on the role of social settings in the accuracy of judgment and decisions is one of the main goals of the current thesis.

Although different in their causal explanations for why people scrutinize others' answers at a greater of lesser extent, the two latter approaches (i.e., Mugny et al.'s conflict elaboration theory and Sperber et al.'s notion of epistemic vigilance) may be reframed in terms of the dual-process approach to reasoning, judgment and decision. More specifically, the information provided in social contexts might be more or less scrutinized depending on whether or not it triggers the engagement of deliberate (T2) processing. For example, people are more likely to further scrutinize divergent information provided by low status sources (e.g., low competence others, minorities), whereas the same information provided by high status sources (e.g., high competence others, majorities) or information that is coherent with people's own intuitions and beliefs (T1 processing), is more likely to be uncritically accepted.

Building on these three lines of research, the general hypothesis of the present thesis is that decision environments may contribute to correct judgmental biases to the extent that they promote people's awareness of their potentially faulty judgments or strategies, thus triggering the engagement in more elaborate reasoning processes. Specifically, decision environments might be corrective 1) when, in continuous contexts, people receive accurate and timely feedback that reveals the inadequacy of their judgmental strategies or 2) when, in socially enriched contexts, they are exposed to the answers provided by others perceived as unreliable (e.g., low in competence). On the contrary, judgmental biases should persist, eventually being reinforced, in decision environments that keep people confident about the quality of their intuitive (T1) answers or strategies (this may happen when the feedback provided in continuous contexts is incomplete or misleading, or when people do not find reasons to treat others' answers with suspicion).

In the Introduction I briefly review research on the HB tradition, from its earlier focus on explaining biases through specific underlying heuristics to the dual-process accounts of reasoning that currently dominate the explanations of judgmental biases in this research field. Then, I will briefly present the abovementioned lines of research (i.e., my inspirational research programs), focusing on the aspects that most influenced the current proposal.

Next, in the three empirical chapters, I present the empirical work developed to test the role that decision environments might play in people's reasoning processes. Each chapter presents a distinctive set of experiments that used different kinds of reasoning tasks to test how 'continuous and socially enriched' decision environments might promote (or hinder) the correction of judgment biases typically reported in HB literature. The chapters were structured as papers for publication (one of them submitted and currently under review, and two in prep.). As a result of this structure, there is some degree of overlap and redundancy of contents in this thesis (both among the chapters and in their relation to Introduction and General Discussion).

In Chapter 2 I present two experiments that used a multiple trial version of the pseudodiagnosticity paradigm (Doherty et al., 1979; Doherty, Schiavo, Tweney, & Mynatt, 1981) to test whether people's tendency to select pseudodiagnostic (i.e., diagnostic worthless) information to test their hypotheses can be reduced (or even reinforced) in continuous learning contexts where people receive feedback that reveal (or mask) their errors.

Chapters 3 and 4 report experimental work that gives special emphasis to the social dimension of decision environments, by testing whether and how being exposed to other people's answers might cue people's engagement in more elaborate reasoning processes. The two experiments included in Chapter 3 used a modified version of the anchoring paradigm

(Strack & Mussweiler, 1997) to test whether people's tendency to perceive others as more biased than themselves (Pronin, 2007; Pronin, Gilovich, & Ross, 2004; Pronin, Lin, & Ross, 2002; Wilson, Houston, Etling, & Brekke, 1996) would make them further scrutinize and adjust away from anchors provided as answers given by a social source (other participants) versus the same anchors presented without a specified source (the modal condition in anchoring research). In Chapter 4 the role that other people's answers might play in triggering more elaborate processing was tested by presenting participants with the answers that other high (vs. low) in competence participant had allegedly provided in several classical reasoning problems that present a conflict between heuristic-based and rule-based (logic or probabilistic) answers and where the modal response for naïve participants are the (T1) heuristic-based answers.

Finally, in Chapter 5 I a) summarize the main findings of this empirical work; b) present their limitations and discuss ways to overcome them; and c) explore their social implications as well as their implications for current dual process theories of reasoning.

Chapter 1 – Introduction

A Historical View on Heuristics and Biases Research

Heuristics and biases early days

The idea that people's judgments do not conform with the prescriptions of the normative rational models was not entirely new when the HB research program was launched (e.g., Edwards, 1968; Meehl, 1954; Simon, 1957). At a theoretical level, Herbert Simon (1955, 1957) had already introduced the concept of "bounded rationality", and claimed that people reason and act rationally, but only within the constraints imposed by their limited computational and informational resources. At an empirical level, when studying how people revised their probabilities in the face of new evidence, Edwards (1968) had also concluded that people reason in accordance with the normative rules of probability (e.g., Bayes' rule), although they underweight the new evidence, thus being conservative in their revisions.

In contrast with these previous proposals, Tversky and Kahneman (1974; Kahneman & Tversky, 1979) argued that the judgmental heuristics that people use are not merely simpler versions of an ideal statistical or logical analysis, rather they entail qualitatively different cognitive processes. Specifically, these authors proposed that people's intuitive strategies of estimating probabilities, frequencies, and other uncertain quantities were dominated by case-specific instead of statistical information. In this context, three heuristics were initially proposed: representativeness, availability, and anchoring-and-adjustment.

The representativeness heuristic consists of replacing the estimation of probabilities by the computation of the similarity between a sample case and the category prototype (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971). Specifically, the perceived likelihood of the sample case depends on its similarity to the category prototype. As the similarity increases the sample case will be perceived as a more likely outcome regardless of actual probabilities. Whenever similarity and probability covary, representativeness provides fairly accurate and easy to compute (intuitive) judgments under uncertainty. However, when other factors that affect probability do not affect similarity, this heuristic is bound to lead to characteristic and systematic biases. The lawyer-engineer problem (Kahneman & Tversky, 1973) is a classic illustration of these biases. In this problem, participants were presented with personality descriptions that were randomly drawn from a sample composed by 100 lawyers and engineers. Participants' task was to estimate the likelihood that the person described was a lawyer or an engineer. In one condition there were 70 lawyers and 30 engineers in the sample, while in the other condition the base rates were inverted (i.e., 30 lawyers and 70 engineers). The personality description was the same in both conditions and it was prototypical (even if just mildly diagnostic) of an engineer. Despite the fact that the base-rates of being a engineer were almost three times higher in the second case, participants made similar estimates in both conditions. In other words, category membership was estimated based on the extent to which the descriptions were representative (similar to the prototype) of a lawyer or an engineer while base-rates (prior probabilities) were neglected.

When using the availability heuristic, people evaluate the probability of an event based on the ease with which it can be imagined or retrieved from memory. The easier it is to imagine or recall instances of the event, the more likely or frequent the event is perceived to be. The opposite is true when the event is harder to imagine or to recall. This heuristic is sound and ecologically valid since more frequent events tend to be better represented in memory and thus more easily retrieved. However, the fact that ease of retrieval can be influenced by many other factors besides actual frequency (e.g., vividness of the information, media exposure, primacy and recency) leads people to overestimate the risks of events that are mediatic or highly dramatic (e.g., terrorist attacks, airplane accidents, probability of having cancer) and underestimate less impressive or salient ones (e.g., probability of dying from diabetes or from a coconut falling on one's head).

The anchoring-and-adjustment was the third heuristic initially proposed by Tversky and Kahneman (1974). According to these authors, when asked to make numerical estimates under uncertainty, people are influenced by any value that is presented in the decision context. Specifically, people use the value presented in the context as a starting point (or an anchor) and then make the necessary upwards or downwards adjustments. Due to insufficient adjustments, final judgments are usually biased towards initially considered values. To illustrate, in one of the classical demonstrations of the use of anchoring and adjustment heuristic, Tversky and Kahneman (1974) asked participants whether the percentage of African countries in the United Nations was lower or higher than a number (i.e., anchor value: 10 vs. 65) that was randomly given by a wheel of fortune. After responding to this comparative question, participants were asked to indicate their absolute estimates of the percentage of African countries in the United Nations. Although the value presented was perceived as random and thus irrelevant for the judgment at hand, participants absolute estimates were biased in the direction of the anchor value (i.e., the mean estimates of participants presented with the low anchor value were significantly lower than those of participants presented with the high anchor). As research on the HB tradition progressed, more heuristics (and biases) were added to the original list (e.g. simulation heuristic, Kahneman & Tversky, 1982b; affect heuristic, Slovic, Finucane, Peters, & MacGregor, 2007).

These heuristics were said to rely on natural assessments. That is, basic cognitive processes (e.g. feature matching, memory retrieval) that the human mind had evolved to make, and for this reason they were expected to yield accurate judgments when used in the decision environments that shaped them through evolution. However, modern human environments have changed so much, and so quickly that heuristics can also give rise to severe and systematic errors (Gilovich et al., 2002; Kahneman & Tversky, 1972; Tversky & Kahneman, 1971, 1973, 1974; for a recent review see Kahneman, 2011). The ecological validity of the heuristics is thus at stake and it is largely an empirical question. The challenge that modern humans face is a metacognitive one. We have to adapt to dynamic and complex decision contexts and decide when heuristics can be trusted and when heuristic-based outputs need to be prevented or overridden.

Interestingly, the early studies of HB research program focused on specifying the conditions under which the use of the heuristics results in errors and biases (i.e., answers that depart from the prescriptions of normative models). The little interest initially displayed in the conditions under which the errors produced by the use of heuristics were prevented or overridden was justified by the assumption that correct answers produced by controlled reasoning were the default and thus need no explaining (Kahneman & Frederick, 2002).

Despite the adequacy of these heuristics in most contexts was never meant to be in question, the accumulation of demonstrations of people's violations of rules of logic and statistics led critics to accuse Kahneman and Tversky of presenting an unfairly negative view of the human intellect (e.g., Cohen, 1979, 1981; Einhorn & Hogarth, 1981; Gigerenzer, 1991).

Regardless, the initial goal of the HB research program was to identify for each heuristic a set of unique biases. For instance, representativeness heuristic was suggested to explain biases of probability judgments like the gamblers' fallacy and non-regressive predictions, people's belief in the law of small numbers, and base-rates neglect (Tversky & Kahneman, 1971). Anchoring-and-adjustment was linked to the explanation of judgmental biases such as preference reversals (Lichtenstein & Slovic, 1971), hindsight bias (Fischhoff, 1975), people's conservatism in using cues to predict a criterion (Lichtenstein, Earle, & Slovic, 1975), people's overestimation of conjunction events and their underestimation of disjunctive events (Sherman & Corty, 1984), and the perseverance of initial impressions (Asch, 1946; Ross, Lepper, & Hubbard, 1975).

Although providing a general understanding of human judgment, this initial proposal was criticized for being too vague and difficult to test in a rigorous way. Heuristics were said to be "mere descriptions after the fact" (i.e., the heuristics were described by the biases they were supposed to explain, and the same biases can be explained by different heuristics), thus having little predictive value (Gigerenzer, 1996; Gigerenzer & Todd, 1999; but see Braga et al., 2018; Sherman & Corty, 1984).

On the other hand, this initial proposal revealed itself to be insufficient to explain when and how case-specific information would be dominant over the statistical one[1]. Not before too long further research started to qualify Kahneman and Tversky original findings.

For instance, Ajzen (1977) showed that base-rates were are no longer neglected when people were given (or could infer) a causal explanation for an event (e.g., a high proportion of failures on an exam is inferred to result from a difficult exam); Bar-Hillel (1979, 1984) showed that people are sometimes responsive to the relative size of a sample when compared to the size of the original population; Tversky and Kahneman (1981) themselves provided evidence showing that biases associated with heuristic processing could be reduced or eliminated by certain framings of problems. These new findings led researchers to broaden the focus of their research from the study of specific aspects of each heuristic (e.g., the substitution of probability judgments by judgments of similarity or ease of retrieval) to the study of heuristic reasoning, and the conditions under which it produces errors vs. accurate judgments (Kahneman & Tversky, 1982a; Tversky & Kahneman, 1983).

The statistical heuristics approach

According to the statistical heuristics approach (Jepson, Krantz, & Nisbett, 1983; Nisbett, Krantz, Jepson, & Kunda, 1983), the prevalence of biases and errors reported in the HB program results from the (high) difficulty inherent to the original problems (see also Evans & Dusoir, 1977). Factors such as a) the clarity of the sample space (i.e., how easy it is to discern what are the possible outcomes) and the sampling process, b) the recognition of the operation of chance factors; and c) cultural prescriptions to use (or not) formal rules, make these problems

[1] As aforementioned, Tversky and Kahneman never claimed that people's reasoning always relied on heuristics, neither that these heuristics always led to errors (Kahneman & Tversky, 1982a). Regardless, at its earlier days, the HB research program was mostly focused on the specific situations under which the use of heuristics resulted in biased judgments and no so much in establishing the boundary conditions that lead people to use or avoid these heuristics (Kahneman & Frederick, 2002).

difficult to respond and lead people to rely on judgmental heuristics instead of statistical heuristics. In support of this perspective, Nisbett et al. (1983; Jepson et al., 1983) showed that statistical answers were more frequent when the problems used involved events with clear sample spaces and sampling processes (e.g., problems about abilities and achievements in contrast with problems about personal relationships) (Jepson et al., 1983, Experiment 1), when the role of chance is made salient (e.g., by explicitly referring to a random device underlying the case selection) (Nisbett et al., 1983, Experiment 3), and when using problems involving domains for which there are cultural prescriptions to use statistical reasoning (e.g., problems about sports) (Nisbett et al., 1983, Experiment 4).

Furthering this research, Fong, Krantz, & Nisbett (1986, Fong & Nisbett, 1991) claimed that statistical heuristics are rudimentary but abstract versions of statistical rules, and showed that people's statistical reasoning improved (in frequency and in quality) with formal instructional methods. While the initial definition of statistical heuristics as intuitive versions of some statistical principles (Nisbett et al., 1983) make them the intuitive statistical counterpart of the judgmental heuristics (i.e., availability and representativeness), the idea is that they pave the way to the development of formally learned rules for reasoning (Fong et al., 1986) making them, in the end, instances of controlled processing.

In sum, by proposing that people possess some level of statistical intuition that can be used to develop through learning rule-governed statistical reasoning, the statistical heuristics approach anticipated the dual-process theories of reasoning and decision-making (e.g., Evans & Stanovich, 2013). According to these theories, human judgment is better conceptualized in terms of largely autonomous heuristic processes (both statistical and non-statistical) and deliberate rule-based reasoning (also both statistical and non-statistical).

Dual-process theories of reasoning

More recently, research in the HB tradition gave rise to dual-process accounts of reasoning, judgment and decision making (e.g., Barbey & Sloman, 2007; Epstein, 1994; Evans, 2008; Evans & Over, 2013; Evans & Stanovich, 2013; Kahneman, 2011; Kahneman & Frederick, 2002; Sloman, 1996, 2014; Stanovich, 1999, 2011, 2018).

All these dual process theories share the assumption that people's judgments are governed by two types of processing that roughly correspond to the classic distinction between intuition (Type 1 processing, hereafter, T1) and reflection (Type 2 processing, hereafter, T2). The distinctive feature of these two types of processing is their dependency on working memory. While T1 is autonomous (i.e., it makes minimal demands on working memory resources), T2 involves cognitive operations (e.g., decoupling and hypothetical thinking) that are strongly dependent on working memory capacity (Evans & Stanovich, 2013).

These theories diverge, however, in how they conceptualize the interaction between T1 and T2. According to the parallel-competitive theories (Denes-Raj & Epstein, 1994; Sloman, 1996), the two types of processing run in parallel and compete with each other. From this perspective, reasoning errors occur when T2 fails to inhibit T1 faulty answers (similarly to what happens with visual illusions). In contrast, the dominant default-interventionist theories (Evans & Stanovich, 2013; Kahneman, 2011) suggest a sequential activation of the two systems. When faced with a reasoning problem, a default T1 answer comes to mind rapidly and with little effort. This answer may be later endorsed, corrected or overridden by the intervention of T2. Errors of reasoning are typically attributed to monitoring failures of T2 in detecting T1 biased outputs.

It is important to note that regardless of the assumption that T2 often has a corrective role, dual process approaches are in general neutral with respect to the relationship between the accuracy of the answers and the type of processing that produced them. In other words, the correctness of an answer is not a defining feature of these two types of processing, which means that T1 and T2 can both produce correct and incorrect answers (e.g., Evans, 2019; Evans &

Stanovich, 2013; Sloman, 1996; Stanovich & Toplak, 2012; Thompson, 2009). However, these features (i.e., being correct or incorrect and result from T1 or T2) are often correlated in the reasoning tasks that present a conflict between T1 and T2 outputs and that are typically used in this research tradition (Evans, 2012; Evans & Stanovich, 2013).

The previously discussed lawyer-engineer problem is a typical example of these reasoning tasks. This problem cues a compelling intuitive T1 answer based on the case specific information (i.e., a stereotypical description of an engineer) that conflicts with the response cued by statistical information (i.e., the base-rates). Although a stereotypical description alone might prompt the T1 answer that the person described is an engineer (i.e., judgment by representativeness), when there are more lawyers than engineers (e.g., 90/10) and the target person is randomly chosen from the total sample, a sampling rule (T2) suggests otherwise. In such cases, to appropriately answer the reasoning problem, people need to engage in T2 processing to second guess the often highly appealing, intuitive (but wrong) T1 answer and override it in order to produce a more accurate, rule-based answer[2].

Therefore, it is crucial to better understand what conditions make people recognize the output of T1 as potentially faulty and trigger the engagement of T2. This question became a topic of interest in the heuristics and biases field, and two types of variables have been found to influence the likelihood of engaging in T2. Namely, 1) features of the decision-maker including cognitive abilities and thinking dispositions (e.g., De Neys, 2006; Stanovich, 1999, 2009) and cognitive impulsiveness (Frederick, 2005), as well as more transient features such as mood (Bless & Schwarz, 1999; Isen, Nygren, & Ashby, 1988) exposure to statistical

[2] It should be noted that the actual normative status of the "correct" answer in this kind of base-rate problems is a matter of debate (e.g., Gerd Gigerenzer et al., 1988). In here we are concerned with the empirical question as to what extent people take the base rates into account during decision making whether or not the base rates ultimately turn out to be ‘‘normative’’ or not (see De Neys & Glumicic, 2008 for a similar point).

thinking (Agnoli, 1991; Agnoli & Krantz, 1989; Nisbett et al., 1983), and metacognitive cues (e.g., answer and perceptual fluency; Alter, Oppenheimer, & Epley, 2013; Alter, Oppenheimer, Epley, & Eyre, 2007; Thompson, 2009; Thompson et al., 2013); and 2) the characteristics of the task: the amount of time allocated to the task (e.g., Evans & Curtis-Holmes, 2005; Finucane et al., 2000), the instructions provided (Daniel & Klaczynski, 2006; Evans, 2002; Newstead, Pollard, Evans, & Allen, 1992; Vadeboncoeur & Markovits, 1999).

Importantly, by focusing attention on these two types of variables, previous research has left largely unexplored the role that the decision environments might play in triggering T2. In this thesis, I take the view that the flexibility of the cognitive system should allow people to adjust the dependence on T1 and T2 according to the decision environments (Ferreira, Garcia-Marques, Sherman, & Sherman, 2006). In other words, whenever the decision environment provides adequate feedback showing the inefficiency of heuristic judgments, this should signal the need to revise T1 autonomous processes and engage in controlled T2 reasoning.

To illustrate, in the case of reasoning tasks such as the lawyer-engineer problem (Kahneman & Tversky, 1973), the extent to which people recognize heuristic reasoning as inefficient should depend on how diagnostic is the feedback provided by the decision environment. In other words, feedback provided on accuracy when responding to problems where T1 and T2 answers coincide (i.e., no-conflict problems, as is the case when base-rates converge with the stereotypical description) should contribute to reinforce the use of heuristic-based answers (allowing people to be right for the wrong reasons). In contrast, the same feedback provided when T1 (biased) answers oppose T2 answers (i.e., conflict problems, as is the case when base-rates oppose the stereotypical description) allow people to be confronted with their own errors providing learning opportunities to second guess T1 wrong intuitions and to more often engage in T2 to better scrutinize problems before answering.

In the following sections I will briefly review literature that explores how judgment unfolds according to the decision environment. Firstly, I will focus on research that tested whether or not the feedback provided in different decision environments contribute to make people aware of the adequacy of their judgmental strategies and, as a consequence, help (or hinder) them to make the necessary adjustments in their judgmental strategies. Then I will turn to research that tested how people adapt their reasoning strategies according to their social contexts. Although reasoning frequently occurs when people are confronted with others, this particular (social) dimension of decision environments has been largely neglected in previous research in the HB tradition.

Human Judgment and the Decision Environment

Although widely accepted that human judgment unfolds in the interaction between the decision-maker and her environment, research in the HB tradition has often neglected the role that the decision environment might play in one's reasoning processes.

This aspect has been often criticized by the opponents of the HB research program with the general argument that heuristics are adapted to natural decision environments, either because they result from years of evolution (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999; Goldstein & Gigerenzer, 2008; Hertwig & Ortmann, 2007; Todd & Gigerenzer, 2003) or because they are progressively adjusted as individuals' judgments unfold in continuous contexts (Hogarth, 1981; Payne, Bettman, & Johnson, 1993). In this context, many judgmental errors and biases reported in the HB research program are claimed to be the result of using experimental procedures that do not preserve the ecology of our natural decision environments, being it from an evolutionary point of view (e.g., the stimuli used are not representative of those that are typically found in the natural environments where humans evolved); or from an individual development perspective (e.g., judgment is studied with discrete inferential tasks whereas most human judgment and decision making is not one-shot but occurs in repeat-playing environments).

In support of the argument that heuristics lead to good quality judgments when the ecology of the environment is preserved, several examples in the literature show that the performance of intuitive strategies is often strongly correlated with the normative ones, which suggests that intuitive strategies are generally accurate (e.g., McKenzie, 1994; Shaklee, 1983; Shaklee & Wasserman, 1986; Wasserman, Dorner, & Kao, 1990). Additionally, some biases typically found in the HB tradition disappear when using experimental tasks that better match those found in real-world contexts. For instance, the overconfidence effect has been found to disappear when people are given representative samples of questions instead of selected samples of questions purposefully chosen to lead to incorrect choices (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin, Winman, & Olsson, 2000; see also recognition heuristic); the proportion of bayesian inferences increases when statistical information is presented in terms of natural frequencies instead of probabilities (Gigerenzer & Hoffrage, 1995) even when participants are experts who make medical and forensic inferences (e.g., Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000, but see Evans, Handley, Perham, Over, & Thompson, 2000; Sloman, Over, Slovak, & Stibel, 2003).

Continuous learning contexts and feedback

In a slightly different vein, Hogarth (1981) argued that the judgmental biases found in the HB tradition are often indicative of cognitive processes that are functional in continuous environments. Rather than using the correct strategy from the beginning (as required by HB one-shot inferential tasks), in continuous environments people can progressively correct their answers by adjusting their judgmental strategies, through repeated trial and error cycles (Hogarth, 1981; see also Brehmer, 1996; Kleinmuntz, 1985).

To illustrate this point, Hogarth compared people's judgments to the process of shooting at a target. If people are given one sole opportunity of hitting the target, it is most likely that they fail. However, the probability of hitting the target increases if people can progressively adjust their shooting strategies (e.g., adjusting the direction, approaching the target), as a consequence of their previous failures. Similarly, human judgment should be better viewed as the result of progressively adjusting one's judgmental strategies instead of choosing a priori the most appropriate one. Remarkably, this idea of judgment as a step-by-step process was part of the definition of the term heuristic originally proposed by the mathematician George Pólya (1887–1985): "a sort of reasoning not regarded as final and strict but as provisional and plausible only, whose purpose is to discover the solution of the present problem" (Pólya as cited in Keren & Teigen, 2004).

It follows from this perspective that judgmental biases found with one-shot tasks may be erased or attenuated provided that people become aware of their errors and have opportunities to make the necessary corrective adjustments later on. In support of this idea, results from few studies that used classical HB reasoning problems with repeated trials showed that providing participants with more opportunities to answer similar problems decreases base-rates neglect (e.g., Fischhoff, Slovic, & Lichtenstein, 1979; Harrison, 1994; Hinsz, Tindale, Nagao, Davis, & Robertson, 1988; Lindeman, van den Brink, & Hoogstraten, 1988; Lopes, 1987; Manis, Dovalina, Avis, & Cardoze, 1980), preference reversals (e.g., Berg, Dickhaut, & McCabe, 1995; Chu & Chu, 1990), and pseudodiagnostic choices (Doherty et al., 1981). We should note that, in some of these experiments, participants' performance increased even when no experimental feedback was provided (e.g., Fischhoff et al., 1979; Harrison, 1994; Hinsz et al., 1988), suggesting that practice, per se, helped participants to adjust their judgmental strategies. However, when feedback was provided, participants' performance became even closer to the normative answers (e.g., Doherty et al., 1981; Lindeman et al., 1988; Lopes 1987; Manis et al., 1980).

Taken together, these results match the commonly held assumption that feedback might help to correct erroneous judgments by making people aware of their errors. In other words, feedback might help people to recognize the insufficiency of their T1 answers, and thus trigger the intervention of T2.

However, feedback might not always be corrective. Our decision environments are full of examples in which feedback is scarce, incomplete, or misleading. In this type of environments, learning about one's judgmental accuracy is difficult, which may lead people to create and maintain erroneous beliefs (Einhorn, 1980; Einhorn & Hogarth, 1978; Hogarth, 2001). The illusion of validity (i.e., people's overconfidence in their erroneous judgments; Einhorn & Hogarth, 1978) is a clear example of these erroneous beliefs. To illustrate, think about the process of recruiting job applicants. When evaluating the quality of their hiring decisions, managers have access to the performance of the accepted applicants, but they know nothing about the performance of the rejected ones. Notwithstanding the limited value of this incomplete feedback, the success of the accepted applicants is taken as proof of the good quality of the hiring decisions, and increases managers' confidence in their decisions. Indeed, if the rejected applicants were also successful, the quality of the hiring process would be low (Einhorn & Hogarth, 1978).

Kind and wicked environments

Extending the idea that feedback may not always be corrective, Hogarth (2001, 2010, see also Hogarth et al., 2015; Hogarth & Soyer, 2011; Soyer & Hogarth, 2015) introduced the concepts of *kind* and *wicked* environments. In kind environments, people receive timely, complete, and accurate feedback on their judgments. This feedback can be taken at its face value as it correctly represents the situation (i.e., negative feedback signals erroneous answers that should be corrected, while positive feedback signals correct answers to be maintained). Weather forecasts are commonly cited as an example of judgments that take place in kind environments, in that accurate and timely feedback is received for each prediction.

In contrast, the feedback received in wicked environments is often incomplete, missing, or systematically biased, making these decision environments unfavorable for turning people aware of their errors or inappropriate judgment strategies. The concept of wicked environment is well illustrated with the example of a physician from the early 20th century who became famous for his ability to diagnose typhoid fever in its early stages. Given that the appearance of the tongue was considered highly diagnostic at that time, the physician's technique consisted on palpating patients' tongues, to later conclude that the patients were infected. The missing point in the physician's analysis was that his technique was actually spreading the disease, instead of helping to diagnose his patients. He was right in his diagnoses, but unfortunately for the wrong reasons (Hogarth, 2001).

Superstitions are also a good example of how wicked environments might prevent people to be aware of their erroneous judgments. Superstitious beliefs consist on a presumed association between two variables, for instance, that taking an exam with a lucky pen will lead to a higher performance. To test the validity of this association, people would need to compare the high vs. low levels of performance obtained when the lucky pen was used vs. not used. However, the fear of a bad performance lead people to avoid using a different pen. Although the feedback they receive from the environment is incomplete (they know nothing about what would be their performance if they had use a different pen), the superstitious belief will be hold as long as the performance in the exam is good.

These two examples illustrate the point that feedback received in wicked environments should not be taken at its face value, as it is influenced by other factors besides the accuracy of the underlying judgmental strategy. Depending on the characteristics of the decision environments (e.g., levels of base rates or selection ratios, the inability to see outcomes of actions not selected), positive feedback might reinforce the use of inadequate judgmental strategies (Einhorn & Hogarth, 1978; Hogarth, 2001; see also Sherman, 1980, on the self-erasing nature of errors). The consequences of this incomplete or misleading feedback are particularly pernicious given people's little motivation to question how successes are achieved. As long as they find a strategy that works to solve a set of problems, it is unlikely that people explore other alternatives, thus failing to notice that much simpler strategies or rules could also solve the same problems (Luchins as cited in Hogarth, 2001; Schwartz, 1982). This reluctance to seek alternative and possibly more effective strategies is even more pronounced when the learned connections were reinforced by external rewards (see Schwartz, 1982 on reinforcement-induced behavioral stereotypy).

The focus of Hogarth's research program (1981, 2001; Hogarth et al., 2015; Hogarth & Soyer, 2011; Soyer & Hogarth, 2015) was on inquiring how these two types of decision environments determine the quality of what people learn across time. From this perspective, the judgmental biases found in the HB research program might often reflect the process of learning from experience in wicked environments. Rather than focusing on how decision environments shape (through experience) people's intuitive decision processes in the long-run, in this thesis I build on Hogarth's concepts of kind and wicked environments to test whether these decision environments facilitate (or difficult) the detection and override of often biased T1 autonomous processes by signaling the need to engage on T2 reasoning.

Social decision environments

Although a huge part of our decisions is made in social contexts and in the presence of others (real or imagined), most previous research on heuristics and biases has treated human judgment as an individual process, neglecting the role that the social context might have on one's reasoning processes (Larrick, 2016). Indeed, previous research on group decision making focused on how different social phenomena and group dynamics (e.g., conformity, group polarization, groupthink) impacted reasoning (e.g., Postmes, Spears, & Cihangir, 2001; Stoner as cited in Plous, 1993). Other research has explored the social nature of reasoning through the contents used in the reasoning problems (e.g., presenting the classical Wason task with contents for which people's background knowledge about the social world are helpful; e.g., Cox & Griggs, 1982; Dawson, Gilovich, & Regan, 2002; Griggs & Cox, 1982), or by emphasizing the conversational processes on human reasoning (Hilton, 1995; Schwarz, 2014). However, for the most part, all these previous research efforts did not systematically investigate how others' judgments and decisions may influence one's own reasoning.

The present thesis contributes to fill this gap by studying the impact that social contexts might play in people's reasoning processes. Specifically, its focus is on how an individual's judgment and decision making may be influenced by others' responses presented in the decision-making context. At this point, I should also distinguish the present proposal from previous research on the effects of accountability on individuals' reasoning processes (e.g., Lerner & Tetlock, 1999; Simonson & Nye, 1992). While this previous research focused on how the need to justify one's judgments and decisions impacts one's reasoning, in the present thesis I am interested in testing whether people use others' answers to guide their own reasoning (either in the right or wrong direction), and to what extent these others' answers might help (or hinder) people to recognize their T1 answers as potentially faulty, thus triggering the engagement of T2.

As aforementioned my conceptualization of this issue was considerably influenced by two research approaches: the conflict elaboration theory of social influence (Mugny et al., 1995; Pérez & Mugny, 1993, 1996), and epistemic vigilance approach (Mercier & Sperber, 2011, 2017; Sperber et al., 2010). In the following, I will critically review these two approaches in light of my current proposal.

Conflict elaboration theory of social influence

More than half a century of social influence and persuasion research has shown that the extent to which people are influenced by information provided by others depends both on the source of the message, and on its content (see Chanthika, 2006; Petty & Cacioppo, 1984; 1986).

At the earlier days, conclusions from this research were that high status sources (high competence, credibility, power, majorities) are more influential than low status ones (e.g., Deutsch & Gerard, 1955; French & Raven, 1959; Hovland, Janis, & Kelley, 1953). As research progressed, it has been proposed that these two types of sources exert different types of influence and induce different kinds of thinking processes. Majorities (or high status sources) are said to induce conformity and convergent thinking and their influence occurs at manifest level (Nemeth, Mosier, & Chiles, 1992; Nemeth & Wachtler, 1983). Given that majorities are perceived as competent sources, their views are accepted as evidence about reality (i.e., informational influence, Deutsch and Gerard, 1955). On the contrary, minorities are said to induce divergent thinking, and their influence occurs at latent or private level (Moscovici, 1980; Nemeth, 1986; Nemeth & Kwan, 1985, 1987).

Summing up, exposure to opposing views emanating from a minority leads to divergent thinking, a process that involves a consideration of the problem from varied viewpoints. Overall, such influence tends to aid performance. On the other hand, exposure to opposing majority viewpoints leads to convergent thought. That is to say, individuals concentrate on the proposed reasoning and answers to the exclusion of other alternative considerations and points of view. Overall, this form of influence does not help performance and may end up to be an impediment (Nemeth, 1986).

Furthering this previous research, the conflict elaboration theory (Mugny et al., 1995; Pérez & Mugny, 1993, 1996) provided a particular insightful framework to study social influence on reasoning processes by claiming that social influence dynamics vary according to the nature of the task at stake.

A fundamental assumption of this theory is that any social influence process is a consequence of a divergence between the source and the target of the information. This divergence gives rise to a conflict, and the meaning that people attribute to this conflict (i.e., conflict elaboration/regulation) is the basic explanatory mechanism through which social influence operates. Specifically, the exposure to information divergent from that held by a target person creates a conflict that leads the target to questioning the validity of her divergent answers. The way people solve this conflict determines the extent of the social influence and is dependent on both the source of the answers and the type of tasks in question (see Doise & Mugny, 1979; Doise, Mugny, & Pérez, 1998; Doise, Mugny, & Perret-Clermont, 1975; Mugny & Doise, 1978, for a similar perspective on developmental social psychology research).

Within this framework, tasks can be distinguished by the extent to which they allow (or not) for objectively right or wrong answers, and by the extent to which they have some sort of social implication for the decision-maker (e.g., people's answer will define membership to a social category or a social group).

Four categories of tasks result from crossing the abovementioned characteristics: i) objective and unambiguous tasks, ii) aptitude tasks, iii) opinion tasks, and iv) non-implicating tasks (see Table 1, adapted from Mugny et al., 1995). The reasoning problems often used in the HB research tradition fall in the "aptitude tasks" category. These tasks are perceived as having an objective right question (which is not known from the beginning) and they are considered as socially-anchoring tasks in that people's answers allow to rank them in terms of their aptitudes or capacities. As their aptitude is at stake, in this type of tasks people are motivated to seek the correct answer, and the status of the source (e.g., majority vs. minority, high vs. low competence, expert vs. naïve) of the divergent information provided is crucial for its acceptance or rejection. Specifically, when confronted with divergent answers provided by a high-status source (e.g., a majority, a high competence source, an expert) people take for granted that the opposing answer is more valid than their own, and therefore adopt or imitate it with little processing of their own (Butera et al., 1992, 1996; Legrenzi et al., 1991; Nemeth, 1986)

Table 1.
Factors influencing conflict elaboration strategies adopted: existence of objective answers for the problem/task and the social implications of the task.

		Social implications of the task	
		No	Yes
Objective answers	No	Non-implicating tasks	Opinion tasks
	Yes	Objective and unambiguous tasks	Aptitude tasks

On the contrary, divergent answers provided by a low status source (e.g., a minority, a low competence source, a naïve person) lead to a different kind of conflict: on one hand, people are not certain about the correctness of their own answer, on the other hand, they are reluctant to accept the source's answer as the likelihood of the source being correct is perceived as low (Kruglanski, 1989). This so-called 'conflict of incompetences' is predicted to lead to a deeper processing of the task, which eventually results in better performance.

Empirical evidence for this theory comes from experiments on inductive reasoning within social settings (Butera, Caverni, & Rossi, 2005; Butera et al., 1996; Butera & Mugny, 1995, 2001; Legrenzi et al., 1991) using a variation of Wason's (1960) rule discovery paradigm. In this paradigm participants are asked to discover the rule underlying a triad of numbers (e.g., 2-4-6). For that, participants are requested to formulate a hypothesis and to propose a triad to test it, after being informed about a possible solution that have been proposed by previous participants in the study. These previous participants were presented as a high- (vs. low-) status source (e.g., majority vs. minority, high vs low competence), and their triads were either confirmatory or disconfirmatory of their hypotheses. Although disconfirmatory testing is more diagnostic, the classical finding within this research paradigm was that people most frequently use confirmatory strategies to test their hypotheses (e.g., Kareev, Halberstadt, & Shafir, 1993; Klayman & Ha, 1987; Rossi, Caverni, & Girotto, 2001; Wason, 1960). Results from Butera and colleagues' experiments showed that people more often followed the source's hypothesis when it came from a majority (a high-status source), and tended to formulate new alternative hypotheses, when it came from a minority (low-status source).

Although participants mainly used a confirmatory strategy for testing their hypotheses (both when source's triad was confirmatory or disconfirmatory), the exposure to the minority's confirmatory strategy (low-status source) increased the use of disconfirmatory strategy. On the contrary, when the majority used confirmation, participants kept using this strategy and almost never used disconfirmation.

Attesting the robustness of these findings, the benefits of being presented with a low- (vs. high-) status source were found both when the competence of the source was directly manipulated (for instance contrasting experts/high competence sources and novices/low competence sources; e.g., Butera et al., 2005; Butera & Mugny, 1995; Maggi, Butera, & Mugny, 1996) or inferred on the basis of its minority vs. majority status (e.g., Butera et al., 1996; Legrenzi et al., 1991). Furthermore, these results were found not only using the Wason's rule discover task but also using other tasks such as information selection tasks (Maggi, Butera, Legrenzi, & Mugny, 1998) and the estimation of lengths tasks (Maggi et al., 1996).

In sum, the conflict elaboration theory predicted and found that being presented with others' divergent answers in aptitude tasks might trigger the engagement of deeper processing of the task (i.e., Type 2 processing) only when these answers come from a low-status source. When divergent answers are given by a high-status source, they will be uncritically adopted without further processing.

Epistemic vigilance research

In contrast with the dominant view of reasoning as a tool to enhance individual cognition (e.g., Evans & Over, 2013; Gilbert, 2002; Kahneman, 2003; Sloman, 1996), Mercier and Sperber (2011, 2017; Mercier & Landemore, 2012; Sperber et al., 2010) claimed that the main function of reasoning is argumentative. In their view, reasoning evolved to allow people to create arguments to convince others, and to evaluate others' arguments in order to avoid being inappropriately convinced or misinformed. Although the full scope of Mercier and Sperber's argumentative theory of reasoning is beyond the scope of this thesis[3], some aspects of it are worth considering here as they inspired my own conceptualization of how reasoning about others' reasoning might affect one's own reasoning.

A central idea of this theory is that people are vigilant towards the information provided by others to avoid being accidentally or intentionally misinformed by them. This so-called epistemic vigilance relies on two different psychological mechanisms: the trust calibration mechanism, targeted at evaluating the trustworthiness of the communicator (i.e., is the source of the information competent and/or benevolent?), and the coherence checking mechanism targeted at evaluating the coherence of the information provided with one's prior beliefs (i.e., is the information provided coherent with previous beliefs?).

[3] The interested reader is referred to Sperber et al. (2010), Mercier and Sperber (2011, 2017).

When presented with information that contradicts their previous beliefs, people either reject the information or revise their previous beliefs. The extent to which information incoherent with their prior beliefs will be rejected strongly depends on how people evaluate its source. In other words, information incoherent with prior beliefs should be more easily accepted when it comes from a highly trusted source (i.e., when it is competent and/or benevolent). On the contrary, when people distrust the source it is more likely that they reject the information provided, and maintain their prior beliefs. On the other hand, when the information provided by others is coherent with one's prior beliefs, it should be accepted with little further scrutiny. The source of the information is, in this case, less relevant to determine its acceptance.

The idea that people are naturally vigilant towards the information provided by others suggests that reasoning failures and judgmental biases typically found with socially decontextualized tasks (as has been the case for most research in the HB tradition) should decrease when these tasks are presented embedded in social settings. In fact, this is a prediction from the theory: "reasoning should produce its best results when used in argumentative contexts" (Mercier & Sperber, 2011, p. 62). In support of this prediction, Mercier and Sperber (2011) refer to research showing that performance in reasoning tasks (e.g. Wason selection task) dramatically increases when participants solve the tasks in groups and are allowed to discuss their answers (Moshman & Geil, 1998; Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006 see also Augustinova, 2008; Bonner, Baumann, & Dalal, 2002; Laughlin & Ellis, 1986; Maciejovsky & Budescu, 2007; Stasson, Kameda, Parks, Zimmerman, & Davis, 1991; but see Argote et al., 1986; Mowen & Gentry, 1980; Tindale, 1993 for counter examples).

More relevant for the purpose of the present thesis, Mata, Fiedler, Ferreira, and Almeida (2013) showed that participants' disposition to be vigilant towards the information provided by others make them adopt a critical mindset when evaluating these others' answers. As a consequence, participants who were more vigilant towards others' answers were better at correctly rejecting their answers and providing the correct response.

Remarkably, in these experiments epistemic vigilance towards others' answers was not triggered by specific characteristics of the source but by participants' bias blind spot (Ehrlinger, Gilovich, & Ross, 2005; Pronin, 2007; Pronin et al., 2002, 2004). That is, participants who believed that they were less intuitive (and, in that context, less prone to biases) than others were more likely to further scrutinize others' answers and detect flaws in their reasoning.

More generally, research on people's naïve theories of bias has shown that people can easily recognize the existence and the impact of most cognitive and social biases in other people's judgments while failing to recognize the role of these biases in their own judgments (Ehrlinger et al., 2005; Pronin, 2007; Pronin et al., 2002, 2004). Such blindness to one's own biases has been said to difficult error correction and, eventually, justify why forewarning people about the effects of various biases have only achieved limited success (Ehrlinger et al., 2005; Pronin, 2007; Wilson & Brekke, 1994; Wilson, Houston, Etling, & Brekke, 1996). However, as Mata et al.'s (2013) experiments suggest, when the reasoning tasks are presented in a socially contextualized format, perceiving others as more prone to biases than the self might make people more reluctant to uncritically follow these others' answers, and trigger the engagement of Type 2 processing.

Overall, the argumentative reasoning theory allows to make predictions similar to those of the conflict elaboration theory of social influence. Specifically, the extent to which people will use other people's answers should depend both on the coherence between one's own answers and others' answers and on the trustworthiness of the source (i.e., her competence and/or benevolence). In other words, others' answers should be accepted with little further scrutiny when they are similar to one's own answers or when they are provided by a high-status source. In contrast, second guessing others' answers and engaging in deliberate processing (T2) should occur mostly when dissimilar answers are provided by a low-status source.

Summary of Current Research

In the next chapters I present the empirical work developed to test the role that decision environments might play in people's reasoning and judgment. As aforementioned, the overarching idea is that decision environments may contribute to correct the judgmental biases commonly found in the HB field to the extent that they promote people's awareness of their faulty judgments or strategies. Specifically, decision environments might be corrective if they help people to recognize their errors or if they cue people to doubt their own judgment, and, as a consequence, trigger the search for alternative strategies through the engagement in more elaborate (T2) reasoning processes. However, by keeping people unaware of their erroneous beliefs and judgments, decision environments might make it difficult to avoid and even reinforce the use of inadequate judgmental strategies.

Three sets of experiments were designed to test this general hypothesis through the manipulation of different aspects of the decision environments. All these experiments used multiple trials versions of experimental paradigms typically used in the HB tradition. In other words, participants answered sequentially to different reasoning problems and received some sort of environmental feedback (either social or not).

The experiments included in Chapter 2 were inspired by Hogarth's (1981, 2001, see also Hogarth et al., 2015; Hogarth & Soyer, 2011; Soyer & Hogarth, 2015) idea that apparently dysfunctional strategies found in one-shot tasks may become adaptive in decision environments that provide people with adequate opportunities for learning (i.e., when accurate and timely feedback provided in kind environments allow people to learn from their errors). On the contrary, these strategies will remain dysfunctional when decision environments do not allow people to learn about the inadequacy of their strategies (i.e., when, in wicked environments, no feedback is provided or when it is incomplete or misleading). These predictions were tested using a multiple trial version of the classical pseudodiagnosticity task (Doherty et al., 1979, 1981). Previous research within this paradigm has shown that people has a robust tendency to select pseudodiagnostic (i.e., diagnostically worthless) information even when diagnostic information is equally available (e.g., Feeney, Evans, & Venn, 2008; Mynatt, Doherty, & Dragan, 1993).

In two experiments, I tested whether this robust tendency can be attenuated (or reinforced) depending on the (kind vs. wicked) structure of the decision environment. In these experiments, participants were asked to make several medical diagnoses based on a limited amount of information that they could ask for. The manipulation consisted on structuring the decision environment so that it clearly revealed the (in)adequacy of participants' choices (kind environment) or it masked the inappropriateness of their choices by making pseudodiagnostic and diagnostic strategies to converge in the same conclusions (wicked environment).

The use of pseudodiagnostic strategies was expected to decrease in kind environment, where making the diagnostic choice was the only strategy that could lead to a correct diagnosis. On the contrary, by allowing people to make a correct diagnosis regardless of making diagnostic or non-diagnostic choices, the wicked environment should hinder the reduction of pseudodiagnostic choices. To test whether the effects of answering several trials in a kind (vs. wicked) environment would transfer to new problems, Experiment 2 of this chapter included a final trial with a new case to diagnose. This problem shared the same deep structure of the previous ones, although it was a more complex version of the original pseudodiagosticity task.

Chapters 3 and 4 give a special emphasis to the social dimension of decision environments, by presenting research designed to test whether and how other people's answers might cue people's engagement in more elaborate (T2) reasoning processes. The ideas for these experiments were derived both from research on the conflict elaboration theory of social influence and from research on epistemic vigilance. As aforesaid, these two lines of research differ in their causal explanations for the effects that social contexts might have in people's reasoning processes. However, they converge on the idea that the information provided in social contexts might be uncritically accepted or further scrutinized depending on 1) how (dis)similar it is to people's background knowledge and beliefs; and 2) what is the status of its source (i.e., whether it is a competent or incompetent).

The two experiments included in Chapter 3 used a modified version of the anchoring paradigm (Strack & Mussweiler, 1997) to test whether the robust anchoring effect (i.e., the assimilation of a numeric estimate to a previously considered value used as an anchor) could be reduced (or reinforced) when participants were forewarned (or not) about the anchoring bias, and the anchor values were presented as other people's answers (or without a specified source). In both experiments, participants were asked to answer several general knowledge questions after considering possible answers allegedly given by a previous participant or presented without a specified source. Before answering these questions, participants were either forewarned or not about the anchoring effects.

The rationale for these experiments was the idea that people should be more critical (and therefore more prone to engage in more deliberate reasoning) when they have reasons to be suspicious about the answers provided. Given that people often recognize the existence and the impact of biases in other people's judgments while failing to recognize the role of these biases in their own judgments (Pronin et al., 2002, 2004), it is proposed that forewarnings about the anchoring bias should make participants perceive other participants, but not themselves, as biased. Therefore, participants should engage in deliberate adjustment, thus reducing the anchoring effect, when they are forewarned about the anchoring bias, and the anchors are presented as other participant's answers. Given that adjusting away from other participant's answers is a corrective process that takes time and effort, participants who were forewarned about the anchoring effect, and then received other people's answers as anchors, were predicted to take longer to give their own responses.

Finally, the two experiments included in Chapter 4 tested the role that other people's answers might play in triggering Type 2 processing, by directly manipulating their competence (Experiments 1 and 2) and the (dis)similarity between these answers and those that most people give in the same tasks (Experiment 2). Specifically, participants were asked to solve several classic reasoning tasks that present a conflict between heuristic-based and rule-base answers. Before responding to each task participants were presented with an answer allegedly provided by other participants in the study. In both experiments, the competence of other participants was manipulated by using experimentally manipulated profiles of people stereotypically high (vs. low) on analytical competence. The answers provided by others were either the heuristic but incorrect answers that most people give to this kind of reasoning problems (Experiment 1) or a mixed set of heuristic and rule-based answers (Experiment 2). To test whether the cognitive effects of reasoning about others' reasoning spill over to subsequent reasoning problems, both experiments included a second set of problems presented without answers from other participants.

The initial prediction was that participants would be more critical towards heuristic answers provided by a low competent other, and would more frequently follow these answers when provided by a high competent other. As a consequence, participants were expected to perform better when answering in the low competence than in the high competence condition because they should more frequently detect low-competent others' incorrect solutions (Experiment 1). Furthermore, when at least part of the answers provided by others are logic or probabilistic valid, and, as such, diverge from the heuristic-based answers that most people usually give, coherence checking mechanisms are expected to prompt further scrutiny of these answers and lead to better performance particularly when these answers come from low-competent others (Experiment 2). The extent to which the performance advantage in the low competence condition was generalizable beyond the initial set of problems was evaluated looking at participants performance in the second set of problems (where answers from others were no longer are provided).

Chapter 2 - Pseudodiagnosticity in Continuous Learning Contexts: The Role of Wicked and Kind Decision Environments

Imagine you are a doctor trying to diagnose which of two diseases affects a patient. Your patient has fever and is covered with a rash. You already know that 84% of people with Disease A have fever. If you could only ask for one more piece of information, which of the following options would you choose: a) the percentage of people with Disease B that has fever, b) the percentage of people with Disease A that is covered with a rash, or c) the percentage of people with Disease B that is covered with a rash?

From a Bayesian perspective, you should select information that allows to discriminate between the two diseases (i.e., that allows to compare the probability of a given symptom across both diseases). In other words, the probability of having fever in Disease B (option a) is the only option that provides you with useful information to make a differential diagnostic of the diseases. However, previous research with this kind of scenarios have shown that people prefer to know more about one single disease (option b), ending up with information that is worthless for the judgment at hand (Feeney, Evans, & Clibbens, 2000; Kern & Doherty, 1982).

This apparent bias to select and use non-diagnostic information when diagnostic data is equally available has been dubbed pseudodiagnosticity (Doherty et al., 1979). Extant research on pseudodiagnosticity has shown its robustness. Pseudodiagnostic preferences have been repeatedly found to dominate hypothesis testing across different domains (Maggi et al., 1998; Mynatt et al., 1993; Van Wallendael, 1995; Wolf, Gruppen, & Billi, 1988), using different materials and presentation formats (Beyth-Marom & Fischhoff, 1983; Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996; Doherty et al., 1981), and even among experts (Gruppen, Wolf, & Billi, 1991; Kern & Doherty, 1982).

Several explanations have been proposed for this tendency, mainly focusing on people's wrong intuitions or limited cognitive resources: (1) people wrongly believe that P(D/H) and P(D/~H) are complementary, which makes unnecessary to ask for these two probabilities (Doherty et al., 1996); (2) when the probability of a data in a given hypothesis (P(D/H)) is high, people wrongly assume that the probability of the same datum for the alternative hypothesis (P(D/~H)) will likely be lower (Doherty et al., 1996; Mynatt et al., 1993); (3) people have a hard time in dealing with numerical information, so that the format in which the information has been presented induce the effect (Beyth-Marom & Fischhoff, 1983; Doherty et al., 1996); and (4) people focus their attention in one specific hypothesis which is less demanding for memory (Girotto, Evans, & Legrenzi as cited in Evans, Venn, & Feeney, 2002; Legrenzi, Girotto, & Johnson-Laird, 1993; Mynatt et al., 1993).

However, by focusing on people's answers to one-shot tasks, most of this previous research has neglected how people's selection of information for hypotheses testing unfolds in continuous environments that offer opportunities to answer to similar problems and eventually receive feedback about previous choices. Going back to our initial example, medical doctors' apprenticeship typically involves long periods of training in adequate environments (e.g., hospital internships), which provide repeated learning opportunities and feedback. The more general point we would like to make is that errors and biases found in one-shot tasks may result from processes that are adaptive and functional in continuous contexts (Einhorn & Hogarth, 1978; Hogarth, 1981, 2001). Our goal in this paper is to test how pseudodiagnostic choices in hypothesis testing typically made in one-shot scenarios may change in decision contexts where participants are presented with a series of similar diagnostic problems, sequentially and with feedback.

Pseudodiagnosticity in Continuous Contexts

To the best of our knowledge, Doherty et al., (1981) is the sole previous research that tested the extent with which making participants aware of their erroneous decisions could attenuate the use of pseudodiagnostic selection strategies in a continuous context. These authors presented participants with four trials with different versions of decision making scenarios similar to the one used in the original pseudodiagnosticity study (Doherty et al., 1979). Each trial presented a piece of pottery with several features. Participants' task was to discover from which one of two islands did the piece come from. The probability of each feature in the pottery of each island was presented in a table with the four cells concealed. To make their decisions, participants uncovered two of the four cells based on which they decided on the provenance of the pottery. After making their decisions all participants received feedback about the correct decision and half of the participants were further asked to uncover a third cell of the table. By doing so they were exposed to one diagnostic pair of data per trial.

Although there was evidence that feedback and exposure to diagnostic data co-acted to induce participants to change from pseudodiagnostic to diagnostic strategies, participants' data selection strategies were mainly driven by their success. If the selection strategy was followed by the correct decision (signaled by positive feedback) then its use increased in the following trials. In other words, positive feedback alone reinforced the use of both diagnostic and non-diagnostic strategies.

Kind and Wicked Environments

In discussing the conditions that affect learning from experience, Einhorn and Hogarth (1978, 1981; see also Brehmer, 1980; Einhorn, 1980, 1986; Hogarth, 1981) pointed out that feedback is not always corrective and that, depending on the decision environment (e.g., levels of base rates and selection ratios, or the inability to see outcomes of actions not selected), it can even reinforce inappropriate behavior.

Furthering this initial research, Hogarth (2001) distinguished between kind and wicked environments. In kind environments, people receive timely, complete, and accurate feedback on their judgments. This feedback can be taken at face value as it correctly represents the situation (i.e., negative feedback signals erroneous answers that should be corrected, while positive feedback signals correct answers to be maintained). To illustrate what he meant by a kind environment Hogarth referred to the experience of professional tennis players. By playing every day, these professionals receive accurate and immediate feedback not only from their coaches but also from their direct experience of the consequences of their actions (i.e., successful and failed shots). By making the players aware of strengths and weaknesses of their tactics, the feedback received in this kind environment provide them with the necessary conditions for learning and improvement.

On the contrary, the feedback received in wicked environments is often incomplete, missing, or systematically biased, making these decision environments unfavorable for turning people aware of their errors or inappropriate judgment strategies (Hogarth, et al., 2015; Hogarth & Soyer, 2011). The assessment of recruitment processes in organizations is an often-cited example of how misleading can be the feedback received in these wicked environments. When evaluating the quality of their recruitment practices, HR managers receive feedback about the success or failures of the recruited candidates but know nothing about those who were not recruited. Although the success of the recruitment process is often equated with the success of the recruited candidates, this conclusion is unwarranted unless the managers could measure the (in)success on the excluded candidates. In other words, the positive feedback received in this scenario (i.e., the success of the recruited candidates) should not be taken at face value, as it may result from factors other than the quality of the decisions made (in this specific case, it may result from the inability to see outcomes of actions not selected (Einhorn & Hogarth, 1978; Hogarth, 2001).

The misleading nature of feedback received in wicked environments might be particularly pernicious given that people do not usually question how their successes were achieved. In other words, once they find a strategy that works to solve a set of problems, people are reluctant to seek alternative and possibly more effective strategies to solve the same problems (Luchins, as cited in Hogarth, 2001; Schwartz, 1982).

In sum, the above presented decision environments critically differ on the extent to which they allow people to become aware of their errors or inadequate strategies and, as a consequence, reinforce the use of inappropriate rules or strategies (wicked environments) or trigger the search for different and more useful ones (kind environments) (Hogarth et al., 2015; Hogarth & Soyer, 2012). We should note that this idea is aligned with most prediction- and error-based theories of learning, according to which people learn more rapidly when they err than when they make the correct predictions or choices (e.g., Gluck & Bower, 1988; Hohwy, 2013; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972).

Overview of Present Research

Building on this previous research, the present experiments were designed to test whether the use of pseudodiagnostic strategies can be attenuated (or reinforced) depending on the (kind vs. wicked) structure of the decision environment.

In the two experiments here reported, participants answer several medical decision making scenarios similar to those used by Kern and Doherty (1982). In each scenario, participants have to decide which one of two diseases affected a patient who presented two symptoms. Information about the prevalence of these symptoms in the two diseases is presented in a 2 x 2 table with the four cells concealed. To make their diagnosis, participants may ask for two of the four pieces of information in the table: P(Symptom1/Disease1), P(Symptom1/Disease2), P(Symptom2/Disease1) or P(Symptom2/Disease2).

After choosing the two pieces of information participants make their diagnosis decision, and then see the table with all the cells uncovered. Revealing the values in the four cells allow us to manipulate the structure of the decision environment by varying the values presented in cells b and c[4].

These values either correspond (wicked environment) or not (kind environment) to what a pseudodiagnostic reasoner would intuitively expect to find in these cells (Evans et al., 2002; Doherty et al., 1996; Mynatt et al., 1993). Specifically, a wicked environment was created by making the percentage of cases in cell b always smaller than that of cell a; and/or by making the percentage of cases in cell c as great or greater than that of cell a. In a kind environment the percentage of cases in cell b was always larger than that of cell a; and/or the percentage of cases in cell c smaller than that of cell a (see Figure 1).

Wicked environment

Type 1	Disease A	Disease B
Symptom 1	**86%** (cell a)	28% (cell b)
Symptom 2	98% (cell c)	46% (cell d)

Kind environment

Type 1	Disease A	Disease B
Symptom 1	**86%** (cell a)	98% (cell b)
Symptom 2	70% (cell c)	84% (cell d)

Figure 1. Example of critical trials in wicked and kind environments.
Note. The actual presentation of the table information was dynamic. Bold type values correspond to the value presented to participants' in their first choice, independently of what cell they had chosen.

[4] For ease of presentation, "cell a" is here defined as the first cell participants choose, and "cell b", "cell c" and "cell d" correspond to the second cell chosen when participants followed a diagnostic (cell b), a pseudodiagnostic (cell c) or a non-diagnostic (cell d) selection strategy.

The wicked environment provides misleading feedback about the appropriate strategy to select information in that both diagnostic and pseudodiagnostic choices lead to the same conclusion about the diagnosis. For instance, in the scenario presented in Figure 1, participants would conclude that the patient had Disease A, regardless of choosing cell b or cell c. Participants in this condition are thus expected to pursue using pseudodiagnostic strategies throughout the trials. On the contrary, in the kind environment, diagnostic choice is the only strategy that leads to a correct diagnosis (although the patient had Disease B, participants would conclude otherwise if they made the pseudodiagnostic choice), which is expected to reinforce the use of this strategy and reduce pseudodiagnostic choices.

Experiment 1

Method

Participants and design. One-hundred thirty-eight (113 females, M_{age} = 21.67, *SD* = 5.42) volunteers participated in the study. Part of the sample was composed of second-year psychology students (*N* = 98, 84 females, M_{age} = 20.97, *SD* = 4.81) who participated for a credit course. The remaining 40 participants (29 females, M_{age} = 23.40, *SD* = 6.42) were recruited from a pool of participants using ORSEE (Greiner, 2015), and received 10€ for their participation. In both cases, the experiment was part of a 1-hour experimental session that included other studies about memory and social perception. Up to eight participants were recruited for each experimental session.

The experiment followed a 2 (decision environment: kind vs. wicked) x 4 (blocks of trials: B1 – B4) mixed design, with the former factor manipulated between-participants, and the latter manipulated within-participants.

Prior approval for running the experiment (and also Experiment 2 reported in this paper) was obtained from the Ethics Committee of the Faculty of Psychology, University of Lisbon.

Materials. Two decision environments were created with the goal to either reinforce (wicked environment) or to correct (kind environment) pseudodiagnostic choices, depending on the values presented in cells b (diagnostic choice) and c (pseudodiagnostic choice) (see Figure 1).

When a pseudodiagnostic choice (i.e., choosing two pieces of information about the same disease) is made in the wicked environment, the values presented in the two cells (a and c or b and d) are always high and higher than the corresponding values in the alternative disease. Hence, the structure of this environment makes the diagnostic choice (cell b) seem superfluous as it corresponds to what people expect (Evans et al., 2002; Doherty et al., 1996; Mynatt et al., 1993).

Moreover, this environment allows participants to be right for the wrong reasons, since both diagnostic and pseudodiagnostic information coincide in the same conclusion about the diseases that affect the patients. From the viewpoint of prediction- and error-based theories of learning, the structure of this decision environment hinders participants' learning about the inappropriateness of their information selection strategies by i) presenting information that coincide with what they already expected (i.e., there is no discrepancy between their prediction and the actual outcome), and ii) by allowing participants to make the correct diagnoses even when they had selected normatively irrelevant information (e.g., Gluck & Bower, 1988; Hohwy, 2013; Kamin, 1969; Kruschke, 1992; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972).

In contrast, when participants made a pseudodiagnostic choice in kind environments, the values presented in the two cells are also high, but the corresponding values in the alternative disease are even higher. In this case, the structure of the decision environment highlights the relevance of choosing cell b (i.e., making diagnostic choices) and allows participants to discriminate between the outcomes of diagnostic and pseudodiagnostic strategies (since they lead to opposite diagnoses; see Appendices A1.1. and A1.2. for a detailed presentation of the structures of the decision environments).

Besides the four critical trials with the abovementioned structure, in each decision environment participants are presented with four neutral trials. These trials are included for two reasons: to allow for cells b (in kind environment) and c (in wicked environment) to vary across trials (otherwise participants could anticipate that the values in these cells were always high); and to allow the final diagnosis to vary across trials (otherwise it would always correspond to the first disease searched). Thus, in two neutral trials cell b presented high values and cell c presented low values, but these high and low values were aligned with the structure of the decision environment (i.e., diagnostic and pseudodiagnostic choices lead to the same diagnosis in wicked environment, and to opposite diagnosis in kind environment). In the remaining two neutral trials, cells b presented high and c presented low values. In these trials a) the final diagnosis did not correspond to the first disease searched; and b) pseudodiagnostic choices led to incorrect diagnoses in wicked but not in the kind environments.

Procedure. Participants were randomly assigned to one of the experimental conditions: kind and wicked environment. The experiment used a multiple trial version of the pseudodiagnosticity paradigm (Doherty el al., 1979; 1981; Kern & Doherty, 1982). Participants were asked to imagine they were missionary doctors who had recently moved to a remote island to help in diagnosing the diseases that were affecting the local communities. Accordingly, they would be randomly presented with eight medical diagnosis scenarios in which they had to decide which one of two diseases was affecting their patients.

In each trial, participants were informed about the patient's two symptoms. They could then have access to the prevalence of these symptoms in the two diseases, one of which was afflicting the patient. This information was presented in a 2 x 2 table (Symptoms x Diseases) with the four cells concealed. Participants could ask for two pieces of information in the table by pressing the keys correspondent to the cells they wanted to uncover When pressing the key correspondent to the first chosen cell, participants first received information about the prevalence of the selected symptom in the selected disease, and then were asked to select a second cell (see Appendix A1.3. for the detailed instructions and the list of materials).

Based on the information presented in the two selected cells, they were then asked to indicate which one of the two diseases affected their patient. Once the diagnosis was made participants were asked to rate how confident they were in their decision. The information about the remaining two cells in the table was then uncovered and participants were asked a) to make a new diagnosis based on all the information presented in the four cells; and b) to state how confident they were about this second diagnosis (from 1 – *not confident at all*, to 6 – *totally confident*). The main dependent measure was the proportion of diagnostic choices made throughout the blocks of trials. The confidence ratings and the diagnoses made based on two and four cells were collected to keep participants involved with the task. In contrast to Doherty et al. (1981) procedure, no feedback was provided about the accuracy of the diagnoses.

In order to avoid interference from prior knowledge that people might have about symptoms and diseases, participants were informed in the initial instructions that all the symptoms and diseases presented were fictitious, and that the two diseases in each trial were equally frequent in the region. The instructions also presented a detailed explanation of the available choices, emphasizing through examples of possible values (to be found in the four cells) that cells 'a' and 'b' or 'c' and 'd' were not complementary, and that high and low values were possible in both diseases. The experiment started with a practice trial to make participants familiar with the task.

Results

Practice trial. Results from the practice trial showed that diagnostic choices ('cell b') were less frequent $N = 34$) than pseudodiagnostic choices ('cell c') and non-diagnostic choices (‘cell d’) taken together (*N* = 101). However, pseudodiagnostic choices (*N* = 46) were not more frequent than non-diagnostic choices (*N* = 55).

Diagnostic choices across the four trial blocks. The proportion of diagnostic choices made in each two-trials block was calculated after excluding participants with perfect performance in the first block (*N* = 21) (since for these participants there was no room for improvement)[5].

Table 1.
Mean proportions of diagnostic choices (SE) for wicked and kind environments across the four blocks of trials

	B1	B2	B3	B4
Wicked	.19 (.03)	.20 (.05)	.26 (.05)	.33 (.05)
Kind	.24 (.03)	.48 (.05)	.44 (.05)	.45 (.05)

Table 1 shows the mean proportions of diagnostic choices made in each block for both experimental conditions. A 2 x 4 mixed ANOVA with decision environment (kind and wicked) as a between factor and blocks of trials (B1 to B4) as within factor showed two main effects and one interaction. A main effect of decision environment, such that participants in kind environments made more diagnostic choices than in wicked environments, $F(1,115) = 10.81$,

[5] The analysis without excluding participants with perfect performance in the first block of trials revealed the same pattern of results: participants in the kind environment made more diagnostic choices than those in the wicked environment, $F(1,136) = 15.96$, $p < .001$, $\eta_p^2 = .11$, the proportion of diagnostic choices increased across trials, $F(3, 408) = 4.81$, $p = .003$, $\eta_p^2 = .03$, with this increase being higher in kind than in wicked environments, $F(3, 408) = 3.05$, $p = .029$, $\eta_p^2 = .02$.

p = .001, η_p^2 = .09; and a main effect of blocks of trials, showing that the proportion of diagnostic choices increased throughout trials, $F(3, 345) = 9.76$, $p < .001$, η_p^2 = .08. More relevant to our hypothesis an interaction between decision environment and blocks of trials, $F(3, 345) = 3.82$, p = .010, η_p^2 = .03, revealed that the increase in the proportion of diagnostic choices across trials was higher in kind than in wicked environment.

Accuracy of the diagnoses and confidence ratings. As abovementioned, the diagnoses that participants made based on two or four cells and their correspondent confidence ratings were collected just to keep participants involved with the task, they are not informative to the test of our hypothesis. However, they can provide cues on whether participants understood the task and were not answering randomly.

Table 2.

Proportion of correct diagnoses (based on two and four cells) for wicked and kind environments across the four blocks of trials.

			B1	B2	B3	B4
2 cells	Wicked	ND	0.87	0.86	0.91	0.85
		D	0.67	0.92	0.73	0.85
	Kind	ND	0.44	0.40	0.42	0.52
		D	0.00	0.67	0.66	0.75
4 cells	Wicked	ND	0.93	0.96	0.97	0.94
		D	0.63	1.00	0.97	0.98
	Kind	ND	0.95	0.95	0.93	0.97
		D	0.00	0.94	0.96	0.92

Note. ND = Non-diagnostic choice, D = Diagnostic choice

If this is the case, it is reasonable to assume that both the accuracy of the diagnoses, and the correspondent confidence ratings should be higher when based on four than two cells, and regardless the decision environment. Our results followed this pattern, showing that participants correctly understood the task and were focused on its completion (see Tables 2 and 3 for detailed data).

Moreover, the accuracy of the diagnoses based on two cells (either diagnostic or non-diagnostic choices) can be interpreted as a manipulation check measure: while in wicked environments both diagnostic and non-diagnostic choices lead to the same (correct) conclusion about the diagnoses, in kind environments making diagnostic choices is the only way to make correct diagnoses. For this reason, participants' accuracy should be higher in wicked than in kind environments, and the accuracy of the diagnoses made based on diagnostic and non-diagnostic choices should only differ in kind environments (where diagnostic choices are expected to produce more accurate diagnoses). Once more, our results followed this pattern revealing that our manipulation worked properly (see Tables 2 and 3).

Table 3.

Mean confidence (and SD) on the diagnoses (based on two and four cells) for wicked and kind environments across the four blocks of trials.

		B1	B2	B3	B4
Wicked	2 cells	3.52 (1.01)	3.49 (1.10)	3.45 (1.10)	3.48 (1.06)
	4 cells	4.49 (1.06)	4.51 (1.09)	4.56 (1.20)	4.43 (1.26)
Kind	2 cells	3.21 (1.16)	3.34 (0.99)	3.35 (1.04)	3.36 (1.07)
	4 cells	4.80 (0.85)	4.77 (0.92)	4.64 (0.94)	4.75 (1.01)

Discussion

Results from this first experiment revealed that decision environments significantly influenced the extent to which participants learned to select diagnostic information. Aligned with previous findings (Doherty et al., 1981), participants in both decision environments progressively made more diagnostic choices throughout the trials. However, participants in the kind environment outperformed those in the wicked environment. By allowing people to be right for the wrong reasons, wicked environment hindered people's opportunities to learn. In contrast, the kind environment confronted people with the inadequacy of their choices, which led them to progressively adopt more diagnostic strategies of information selection.

This first experiment left unexplored to what extent the advantage shown by participants in kind environments would be transferred to new situations/problems. Doherty et al. (1981) tested the effectiveness of participants' learning by presenting them with the choices allegedly made by others (diagnostic, pseudodiagnostic or diagonal/cell d) in new decision scenarios. In each scenario participants were asked to comment on others' choices. Results revealed that only a small percentage of participants verbalized a correct understanding of the normative principle, even among those participants who had successfully learnt to select diagnostic information. This is consistent with previous findings showing that people can gain an implicit understanding of the task (e.g., learn to make appropriate choices) without improving their ability to verbally communicate it (e.g., Bechara, Damasio, Tranel, & Damasio, 1997; Friedman, 1998; Granberg & Brown, 1995; Granberg & Dorr, 1998). Thus, instead of testing participants' understanding of the task verbally, in Experiment 2 we explore to what extent the performance advantage shown by participants in kind environment transfers to a new decision scenario.

Experiment 2

This study aims to replicate the learning advantage provided by kind environments found in Experiment 1 and to test to what extent can this advantage be transferred to a new and more complex decision scenario.

Method

Participants and design. Ninety-seven participants (51 females, M_{age} = 33.12, *SD* = 10.10) were recruited using Prolific Academic, and received £1.00 for their participation. The experiment was presented as a study about how people make decisions based on a limited amount of information. All participants were native English speakers with at least 18 years old.

Given that the study run online, an attention check was included to improve data quality (Oppenheimer, Meyvis, & Davidenko, 2009). Twelve participants failed the attention check and were therefore excluded from the analyses. The final sample included 85 participants.

The experiment followed a 2 (decision environment: kind vs. wicked) x 4 (blocks of trials: B1 – B4) mixed design, with the latter factor manipulated within-participants.

Materials. The materials and structure of the decision environments were the same used in Experiment 1, except that a larger proportion of critical trials (six in eight) were included in each decision environment. Moreover, the eight trials were presented in a fixed randomized order to assure that all participants were presented with the neutral trials in the same blocks (second and fifth decision scenarios presented) (see Appendices A2.1. and A2.2.). The goal was to reduce experimental noise resulting from having neutral trials (which do not allow for learning) in different blocks for different participants.

Procedure. The procedure was the same of Experiment 1 with the following modifications. The instructions presented a more simplified explanation of the task, without examples of the available choices and the values that could be found in the different cells (see Appendix A2.3. for the detailed instructions). After participants had completed the eight trials in the study phase, a final trial was presented with an additional case to diagnose. Including this final trial allowed us to test whether or not participants were able to transfer diagnostic decision strategies eventually learnt during the study phase to a new diagnose problem with the same deep structure but with some superficial changes, which made it slightly more complex. Specifically, in the final trial the number of symptoms was increased from 2 to 12. Participants were thus presented with a 12 symptoms by 2 diseases table (see Appendix A2.3.). As before, they could only ask for a limited amount of information (i.e., they could ask to uncover 6 of the 12 cells), to make a diagnose. Participants were further told that they would receive information about the chosen cells only after choosing the six cells. In practice, the experiment ended when participants chose the sixth cell on the table, and no information about the cell values was provided.

An attention check (i.e., asking participants to press a specific key to continue the experiment, instead of using the arrow button available from the previous trials) was included between the study phase and the final trial to make sure that participants were following the instructions.

Results.

Practice trial. As in Experiment 1, results from the practice trial showed that diagnostic choices (N = 29) were less frequent than the remaining non-diagnostic choices taken together (N = 56). Pseudodiagnostic choices (N = 29) were in the same number of diagnostic choices and slightly more frequent than non-diagnostic choices (N = 27).

Diagnostic choices across the four blocks of trials. The eight trials were grouped in four blocks of two trials each. The proportion of diagnostic choices made in each block was calculated after excluding participants with perfect performance in the first block (N = 22)[6] (see Table 4).

Table 4.

Mean proportions of diagnostic choices (SE) for wicked and kind environments across the four blocks of trials

	B1	B2	B3	B4
Wicked	.18 (.04)	.41 (.07)	.39 (.07)	.42 (.07)
Kind	.18 (.04)	.48 (.07)	.48 (.08)	.55 (.08)

A 2 x 4 mixed ANOVA with decision environment (kind and wicked) as a between factor and blocks of trials (B1 to B4) as within factor revealed a main effect of block of trials, such that the proportion of diagnostic choices increases throughout trials, $F(3, 183) = 14.73$, $p < .001$, $\eta_p^2 = .19$. No main effect of decision environment nor Decision Environment x Trial blocks interaction were found (Fs < 1). We further compared the proportion of diagnostic choices in B1 and B4 in both decision environments. Although the proportion of diagnostic choices nominally increased more for kind compared to wicked environments, mere task experience seemed to have an effect, as performance increased in both decision environments, t$(61) = 4.52$, $p < .001$, $d = 0.70$) for kind and $t(61) = 3.14$, $p = .003$, $d = 0.50$ for wicked environments.

Transfer of learning. Table 5 shows the number of pairs (i.e., diagnostic choices) chosen in the final test in each decision scenario. Participants' answers were aggregated in non-

[6] As in Experiment 1, the analysis conducted without these exclusions revealed the same pattern of results: the proportion of diagnostic choices increased across trials, $F(3, 249) = 5.26$, $p = .002$, η_p^2 = .06, regardless of the decision environment.

diagnostic choices (when participants who chose zero, one or two pairs of cells) and diagnostic choices (when participants chose three pairs of cells).

Fisher exact test indicates that the proportion of diagnostic choices in the final test was higher in kind (63%) than in wicked environment (39%) (p = .050, unilateral test). In other words, diagnostic decision strategies acquired during the study phase transferred to a new decision scenario in the case of the kind environment but not for the wicked environment.

Furthermore, performance in the study phase (i.e., proportion of diagnostic choices in eight trials) correlates with the number of pairs chosen in the final test in the kind ($r(30) = .66$, $p < .001$) but not in the wicked environment ($r(33) = .17$, $p = .334$). This corroborates that the transfer of learning from the study phase to the final test only occurred for the kind environment.

Table 5.

Number of pairs of cells chosen in the final test in each experimental condition.

Pairs	Wicked	Kind
0	12	6
1	6	3
2	2	2
3	13	19

Accuracy of the diagnoses and confidence ratings. As in Experiment 1, results from these measures might be viewed as an indicator of participants' understanding and commitment to the task, as well as a check of the decision environments manipulation.

As in Experiment 1, in both decision environments, the accuracy of the diagnoses, and the correspondent confidence ratings were higher when based on four than on two cells. For the diagnoses made based on 2 cells, participants' accuracy was higher in wicked than in kind environments. While in wicked environments participants' accuracy was high regardless of the type of information they have chosen, in kind environments participants' accuracy was high only when they have chosen diagnostic information (see Tables 6 and 7 for detailed data).

Table 6.

Proportion of correct diagnoses (based on two and four cells) for wicked and kind environments across the four blocks of trials.

			B1	B2	B3	B4
2 cells	Wicked	ND	0.80	0.79	0.88	0.68
		D	0.50	0.93	0.92	0.82
	Kind	ND	0.35	0.48	0.57	0.59
		D	0.36	0.83	0.83	0.82
4 cells	Wicked	ND	0.93	0.67	0.98	0.55
		D	0.92	0.59	1.00	0.57
	Kind	ND	0.90	0.65	0.83	0.63
		D	1.00	0.62	1.00	0.64

Note. ND = Non-diagnostic choices, D = Diagnostic choices

Table 7.

Mean confidence (and SD) on the diagnoses (based on two and four cells) for wicked and kind environments across the four blocks of trials.

		B1	B2	B3	B4
Wicked	2 cells	2.95 (1.19)	3.30 (1.21)	3.02 (1.14)	3.21 (1.21)
	4 cells	3.83 (1.18)	3.88 (1.17)	3.70 (1.22)	3.34 (1.39)
Kind	2 cells	3.32 (1.16)	3.37 (1.27)	3.10 (1.12)	2.90 (1.17)
	4 cells	4.12 (1.02)	4.05 (1.21)	3.98 (1.09)	3.63 (1.36)

Discussion

This experiment aimed to test to what extent the learning advantage found for kind environment in Experiment 1 would be transferred to a new, and more complex, decision scenario. Following the tendency found in Experiment 1, the proportion of diagnostic choices progressively increased throughout trials in the study phase for both decision environments. However, the expected interaction between the decision environments and the trial blocks was not significant. Comparisons between blocks 1 and 4 revealed significant differences for both wicked and kind environments.

In a final test consisting of a new and more complex decision scenario, which shared the same deep structure of the previous trials, performance was higher for participants answering in the kind environment than those answering in the wicked environment. This suggests that the improvement in performance in kind environment resulted from a deeper learning of the structure of the task during the initial 4 blocks of trials, whereas the observed increase in performance in the wicked environment was the result of more shallow learning. This transfer of learning for kind environment was further supported by the correlations between performance in the study and the test phase, which were significant for kind but not for wicked environment.

General Discussion

Previous research has shown that when testing two competing hypotheses people tend to select and use worthless information despite diagnostic data is equally available. This tendency, known as pseudodiagnosticity, was found to be pervasive across different domains (e.g., Maggi et al., 1998; Mynatt et al., 1993; van Wallendael, 1995; Wolf et al., 1988) and resistant to several attempts to correct it (Doherty et al., 1996; Mynatt et al., 1993).

Building on the idea that many errors and biases found in discrete contexts turn out to be functional when considered in continuous learning contexts (Einhorn & Hogarth, 1978; Hogarth, 1981, 2001), our experiments used a continuous version of the pseudodiagnosticity paradigm to test how this apparently dysfunctional tendency evolve as a function of the decision environment. For that purpose, two decision environments were created, varying in the extent to which they provided participants with more or less opportunities for learning to select diagnostically relevant information.

The distinction between environments that provide appropriate conditions for learning (kind environments), and those in which learning is likely to be misguided (wicked environments) has been previously made by Hogarth (2001, 2010; see also Hogarth et al., 2015; Hogarth & Soyer, 2012). In line with most prediction- and error-based theories of learning, the critical difference between the two environments is the quality of feedback they provide: while in the former case the feedback is accurate and complete, allowing people to learn from it, in the latter case, it is incomplete, missing or misleading, preventing people from learning or even reinforcing the use of inappropriate rules (Gluck & Bower, 1988; Hogarth et al., 2015; Hogarth & Soyer, 2012; Hohwy, 2013; Kamin, 1969; Kruschke, 1992; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972).

Accordingly, in our experiments the structure of the decision environments was manipulated either to clearly reveal the (in)adequacy of participants' choices (kind environment) or to conceal the inappropriateness of their choices by making pseudodiagnostic and diagnostic strategies to converge in the same conclusions (wicked environments). Specifically, in the kind environment, making the diagnostic choice was the only strategy that could lead to a correct diagnosis, which was expected to reduce the use of pseudodiagnostic strategies. On the contrary, in the wicked environment, either diagnostic or pseudodiagnostic choices would lead to the same conclusion about the diagnosis. By allowing people to be right for both the right and wrong reasons (i.e., regardless of making diagnostic or non-diagnostic choices), feedback in wicked environment does not allow to directly discriminate between both strategies, which was predicted to hinder the reduction of pseudodiagnostic choices.

As expected, participants answering in the kind environment outperformed those in the wicked environment. However, the mere exposure to several decision trials with full disclosure of all information (the four cells) in the end of each trial was enough to improve performance. Indeed, participants in both decision environments progressively made more diagnostic choices in both environments and the additional improvement in performance promoted by kind environment was significant in Experiment 1 but not in Experiment 2.

These results are consistent with those of Doherty et al. (1981) in that providing participants with feedback showing the inadequacy of their choices (kind environments) led them to progressively adopt more diagnostic strategies of information selection throughout the trials. Notably, Doherty et al. (1981) ended their paper suggesting that environments that allow people to accidentally reach correct decisions could even reinforce the use of inadequate strategies of information selection. The contrast between wicked and kind environments put forward in our experiments allowed us to empirically test for this suggestion. However, as aforementioned, and in contrast to Doherty et al.'s (1981) prediction, our results showed that even wicked environments reduced the use of pseudodiagnostic choices, though to a less extent.

Doherty et al.'s (1981) tested the effectiveness of learning in their paradigm by asking participants to comment the strategies of information selection that were allegedly followed by other participants in the study. Results showed that only a small percentage of participants were able to correctly verbalize the normative principle underlying the task, even among those participants who had successfully learnt to select diagnostic information. These results were interpreted as evidence that participants had only gained a superficial understanding of the task. Since people can improve their understanding of the task without being able to verbally communicate it (e.g., Bechara et al., 1997; Friedman, 1998; Granberg & Brown, 1995;

Granberg & Dorr, 1998), in Experiment 2 we tested the effectiveness of learning by presenting participants in the end of the task with an additional, slightly more complex decision scenario that shared the same deep structure of the previous ones. Our results showed that gains in response accuracy acquired in the initial trials were transferred to this final scenario when the environment was kind but not when the environment was wicked.

These results indicate that answering multiple trials in kind environments improved participants' deep understanding of the underlying structure of the task, whereas performance improvement in wicked environment was probably shallower (i.e., more linked to the superficial features of the task) and thus did not generalize to a new task with a different superficial structure.

Theoretical and practical implications. Taken together, these findings contribute to improve our understanding of how people select information to test their hypotheses. Previous research has systematically shown that people tend to select non-diagnostic information even in situations in which diagnostic information is equally available. However, most of this research focused on people's answers to one-shot tasks thus hindering our knowledge about how this apparently dysfunctional strategy of information selection unfolds when people are given more opportunities to answer similar problems and become aware of the outcome of their choices. In two experiments, we used a modified version of the pseudodiagnosticity task that allowed participants to respond sequentially to a set of problems with feedback. Results suggest that in such a repeat-playing context, people progressively learn to select diagnostically relevant information to test their hypotheses. More importantly, we found that the extent and depth of this learning depends on the structure of the decision environment, with kind environment producing better results than wicked environments.

These findings provide empirical support to Hogarth's (1981) seminal idea that many biases and dysfunctional strategies found in discrete tasks might be better seen as functional cognitive processes that produce more or less adaptive judgments depending on the decision environment (see also Einhorn & Hogarth, 1978, 1980). It follows that by preventing participants from learning the task and adjust their judgmental strategies accordingly, one-shot experiments often used to study human judgment might be presenting an incomplete view of human judgment and decision making competencies.

Furthermore, the current findings have practical implications for debiasing. Most previous attempts to avoid judgmental biases have focused on changing people's cognitive processes to better fit the tasks (e.g., teaching people to use appropriate rules and principles, clarifying task instructions and materials) while neglecting the role that continuous decision environments might play in judgmental accuracy (e.g., Arkes, 1991; Fischhoff, 1982; Klayman & Brown, 1993; Nisbett et al., 1983; Soll, Milkman, & Payne, 2014). For instance, previous research has tried to reduce pseudodiagnostic choices by clarifying the structure of task (e.g., explaining that cells a and b are not complementary) (e.g., Doherty et al., 1996), presenting materials verbally instead of numerically (e.g., Beyth-Marom & Fischhoff, 1983; Doherty et al., 1996), and making the task less demanding for memory (e.g., Girotto, Evans, & Legrenzi, as cited in Evans et al., 2002; Mynatt et al., 1993).

Improving judgment by changing the decision environment in ways that better fit people's cognitive processes or that encourage the use of better strategies is another approach that has gained popularity (e.g., Hertwig, Hogarth, & Lejarraga, 2018; Hogarth, 2001; Hogarth et al., 2015; Hogarth & Soyer, 2011; Klayman & Brown, 1993; Soll et al., 2014; Thaler & Sunstein, 2008). Our experiments join this growing body of research by showing that some judgmental biases previously found to be robust and pervasive can be avoided or reduced in decision environments that allow people to learn from their errors and provide opportunities to correct them (i.e., kind environments). Unfortunately, many of our decisions take place in environments that make it difficult to learn from (and correct) judgmental errors (i.e., wicked environments).

According to Hogarth (2001; Hogarth et al., 2015), people should learn to seek and deliberately create kind environments. In other words, people should learn to recognize the limits of their experience (e.g., lack of feedback about the non-taken actions), and how to overcome them. For instance, in the previously mentioned example of the recruitment of job candidates, managers should acknowledge the limited value of the feedback they naturally receive, and actively try to complement it with other kinds of information (e.g., trying to think about reasons that might contradict the feedback received or using the professional networking sites like LinkedIn to find out information about the non-selected candidates).

Limitations and future studies. The impact that kind environments might have on people's judgments has been empirically tested in experiments using behavioral simulations (Hogarth et al., 2015). Just like the experimental procedure used here, these simulations let people experience the negative consequences of their decisions, thus helping them to recognize the need to make eventual adjustments in their judgmental strategies. Given that many of our real-world decision environments are wicked, and that deliberately creating kind environments is not always possible, future studies should test the extent with which experiential learning acquired in lab sessions is transferred to real-world settings.

In order to better capture the impact of the different learning environments, our experiments were designed to ensure that participants' answers and underlying reasoning would not be contaminated by their background knowledge. However, previous research has shown that background beliefs about the information provided (e.g., the rarity of the features presented in the selection task) are also a source of pseudodiagnosticity (Feeney, Evans, & Clibbens, 1997; Feeney et al., 2000) and that experience in kind environments might be insufficient to counteract the effects of prior beliefs on people's judgments (Soyer & Hogarth, 2015). To have a more complete picture of the conditions under which pseudodiagnostic strategies of information selection are attenuated (or reinforced), future studies should test the impact of decision environments when people can use their background beliefs to test their hypotheses.

Chapter 3 – Anchoring in a Social Context: How the Possibility of Being Misinformed by Others Impacts One's Judgment[7]

Anchoring effects or the assimilation of a numeric judgment to a previously considered standard value (the anchor) is one of the most remarkable influences in human judgment due to its striking pervasiveness and robustness (for reviews, see Chapman & Johnson, 2002; Furnham & Boo, 2011; Mussweiler, Englich, & Strack, 2004). Anchoring explanations have been used to account for several judgment biases such as hindsight bias (Fischhoff, 1975), preference reversal effects (Lichtenstein & Slovic, 1971), and biased causal attribution (Quattrone, 1982). Moreover, the consequences of falling prey to this bias have been extensively explored in several different applied domains: medical decisions (e.g., Brewer, Chapman, Schwartz, & Bergus, 2007), judicial decisions (e.g., Englich & Mussweiler, 2001; Englich, Mussweiler, & Strack, 2005, 2006), pricing decisions (e.g., Ariely, Loewenstein, & Prelec, 2006; Northcraft & Neale, 1987) and final agreement of negotiation (e.g., Liebert, Smith, Hill, & Keiffer, 1968; Loschelder, Trötschel, Swaab, Friese, & Galinsky, 2016), among others.

The Selective Accessibility Model (SAM; Mussweiler & Strack, 1999; Strack & Mussweiler, 1997; see also Chapman & Johnson, 1994) is perhaps the most consensual

[7] This chapter is based on the paper: Reis, J., Ferreira, M. B., Mata, A., & Garcia-Marques, L. (submitted). Anchoring in a Social Context: How the Possibility of Being Misinformed by Others Impacts One's Judgment.

explanation for anchoring effects found using the standard paradigm (Tversky & Kahneman, 1974). In this paradigm people answer one comparative question (e.g., Is the Mississippi river longer or shorter than 3000 miles) followed by one absolute question (e.g., What is the length of Mississippi river in miles?). According to SAM, when people answer the comparative question, they test the hypothesis that the target value is equal to the provided anchor (e.g., Klayman & Ha, 1987; Wason, 1960), thereby increasing the accessibility of anchor-consistent knowledge (e.g., Higgins, 1996, 1997). This anchor-consistent information is then used to form the final absolute judgment, which results in an answer that is biased towards the anchor. The often-found resistance of anchoring to forewarning and incentives (Chapman & Johnson, 1994, 2002; Epley & Gilovich, 2005; Tversky & Kahneman, 1974) is usually considered to stem from this biased (and largely autonomous) recruitment of anchor-consistent information process.

Research on debiasing suggests that forewarnings may not be sufficient to ensure that correction occurs. When people are forewarned and made aware of a potential bias, the way they try to correct their judgments depends on their naïve theories about their own susceptibility to that bias (Wilson & Brekke, 1994; Wilson et al., 1996).

Importantly, the externally provided anchors used in extant research have been typically described to participants as randomly generated numbers and usually presented without a specified source (Furnham & Boo, 2011; for exceptions, see Dowd, Petrocelli, & Wood, 2014; Glöckner & Englich, 2015; Langeborg & Eriksson, 2016; Meub & Proeger, 2015, 2018; Mussweiler, Strack, & Pfeiffer, 2000). Presenting anchor values as random numbers, although crucial to exclude anchoring explanations based on conversational norms (Grice, 1975; Mussweiler et al., 2004), may inadvertedly lead participants to believe that they cannot be influenced by such arbitrary anchoring values (Wilson & Brekke, 1994; Wilson et al., 1996) and consequently do little or nothing to avoid anchoring effects.

Moreover, providing anchors without a specified source misses the point that in our daily activity anchor values are not presented in a social vacuum but are often suggested by others. The degree to which our own judgments are anchored by such social sources of judgment should depend on naïve theories about others' judgment accuracy. In fact, forewarning manipulations have been successful in convincing participants that people in general, but not themselves, are susceptible to the described biases (Wilson et al., 1996). This so-called 'bias blind spot' (Pronin, 2007; Pronin et al., 2002, 2004) has been shown to lead people not only to avoid relying on others' judgments but also to trigger a search for alternative answers (Mata, Fiedler et al., 2013). It follows that presenting anchor values as estimates of other persons in a modified version of the standard anchoring paradigm might lead participants to adjust away from these "others' estimates" when they are aware of the anchoring effect. In other words, if externally provided anchors are presented as other people's estimates and participants are forewarned of the anchoring effect, they might perceive these estimates as biased and adjust away from them, thus reducing anchoring effects. Furthermore, given that adjusting away from other people's estimates is a deliberate corrective process that takes time and effort, participants who are forewarned about how people often fall prey of anchoring bias, and then receive answers from other people, should take longer to give their absolute answers.

Two experiments were designed to test these hypotheses. In both experiments, participants answered several general knowledge questions after considering a possible answer allegedly given by a previous participant or presented without a specified source. Before answering these questions, participants were either forewarned or not forewarned about the anchoring effects.

In sum, we predict that participants will show substantially less anchoring (relative to the other conditions) when they are provided with others' estimates as anchors, and receive a forewarning about the anchoring effect.

Experiment 1

Method

Participants. To determine the sample size, we followed Mussweiler et al. (2004), according to which 20 participants are sufficient to find the anchoring effect in within-subjects designs (i.e., low vs. high anchors). Given that we have four experimental conditions, eighty participants (74 females; M_{age} = 23.10, *SD* = 4.19) were recruited for this experiment, using our pool of voluntary participants (ORSEE; Greiner, 2015). Three participants were excluded as they left more than a half of the questions unanswered. The experiments reported in this paper were approved by the Ethics Committee of the Faculty of Psychology, University of Lisbon.

Materials. Participants were presented with 14 general knowledge questions. Each of these questions was accompanied by one possible response (the anchor value), which in half of the trials was a low value (low anchor), and in the other half was a high value (high anchor). Following Jacowitz and Kahneman (1995), these anchors corresponded, respectively, to the 15^{th} and 85^{th} percentiles of the response distribution of an independent sample of participants from the same population of the experimental participants (calibration group) (see Appendix B1.1.).

Two versions of material were developed such that the questions presented with a high anchor in one version were presented with a low anchor in the other. In both versions, the order of questions was randomized for each participant. Two additional questions (and the corresponding high and low anchors) were used as warm-up trials.

Procedure. Participants took part in the experiment individually. Upon arrival to the lab, they were randomly assigned to one of four conditions that result from the orthogonal manipulation of forewarnings about the anchoring effect and the source of the anchors provided in the questions: No forewarning and no source (NFNS; no forewarning was presented and anchor values were provided without a specified source), forewarning and no source (FNS; participants were forewarned about the anchoring effect and anchor values were provided without a specified source), no forewarning and source (NFS; no forewarning was presented and anchor values were presented as other participants' estimates), and forewarning and source (FS; participants were forewarned about the anchoring effect and anchor values were presented as other participants' estimates).

Participants were then presented with 14 trials of two sequential questions. In each trial participants were presented with a general knowledge question followed by a possible (but not necessarily correct) answer (the anchor value), and were asked 1) to rate their agreement with the provided answer (from1- *not at all* to 4 - *totally*), and 2) to give their own answer to the presented question. In the source conditions (NFS and FS) these anchors were presented as answers allegedly provided by other participants who had previously participated in the study. In the no source conditions (NFNS and FNS), the same anchors were presented as possible (but not necessarily correct) answers with no specified source. Before the presentation of the general knowledge questions, participants in the forewarning conditions (FNS and FS) were forewarned about the anchoring effect (see Appendices B1.2. and B1.3. for the instructions and an illustration of the trials). Participants' anchoring on the general knowledge questions and the correspondent response times were the main dependent measures in this experiment.

After answering the 14 general knowledge questions, participants were asked to indicate: a) to what extent they believed to have been influenced by the responses provided to the general knowledge questions (from 1 – *not influenced at all*, to 4 – *totally influenced*) (subjective influence); b) how knowledgeable they were about the questions' subject matters (from 1- *none,* to 4 – *very much*) (subjective knowledge); and c) how they perceived their overall performance (from 1 – *very poor*, to 4 – *very good*) (subjective performance). These measures, were included to explore how participants' metacognitive experiences relate with their objective performance in the anchoring task.

Results

Agreement. Except for the FS condition, mean agreement with the anchor was always above the scale mid-point (*M*s = [2.10 - 2.46], 95% CIs = [1.89, 2.68], see Table 1 in Appendix B1.4.) suggesting that when confronted with the provided responses to the general knowledge questions, participants tended to agree with them. A 2 x 2 ANOVA with forewarning (with and without) and source (with and without) as between-subjects factors showed only a marginally significant main effect of forewarning. Forewarning participants about the anchoring effect decreased their level of agreement with the provided anchors, $F(1,73) = 2.97$, $p = .089$, $\eta_p^2 = .04$.

Planned orthogonal comparisons were carried out to test the hypothesis that differences in anchoring were mainly driven by the condition that combines forewarnings with anchors provided with a social source (FS). The first comparison contrasted the FS condition with all other conditions (contrast weights -3, 1, 1, 1); the second comparison contrasted the forewarning condition (FNS) and the source condition (NFS) with the control condition/No forewarning and No source (NFNS) (contrast weights -1, -1, 2); and the third comparison contrasted the FNS with NFS (contrast weights -1, 1). Results from these comparisons showed a marginally significant decrease in agreement for the FS condition compared to all other conditions, $t(73) = 1.81$, $p = .073$, $d = 0.48$. The two remaining comparisons revealed no differences: NFS and FNS vs. NFNS, $t(73) = 1.23$, $p = .222$, $d = 0.30$. and NFS vs. FNS, $t < 1$.

Low and high anchors. Before analyzing participants' responses and calculating the correspondent anchoring scores, we excluded non-interpretable and nonsensical answers (e.g., answering that penicillin was discovered in 18250) were removed from the analysis. Very few (2%) responses were excluded using this criterion. The remaining responses were transformed into z-scores, separately for each question, and answers following low and high anchors were averaged into two composite scores (low and high anchor estimates).

A 2 x 2 x 2 mixed ANOVA with anchor (low and high) as a within-subjects factor, and forewarning (with and without) and source (with and without) as between-subjects factors showed a main effect of anchor, such that mean values for low anchor trials (M = -0.56, 95% CI = [-0.64, -0.48]) were lower than for high anchor trials (M = 0.53, 95% CI = [0.43, 0.63]), $F(1,72) = 237.26$, $p < .001$, $\eta_p^2 = .77$. This main effect was qualified by a significant Anchor x Forewarning interaction, $F(1,72) = 4.14$, $p = .046$, $\eta_p^2 = .05$, showing that the difference between high and low anchors (i.e., the anchoring effect) is larger when participants were not forewarned about the potential biasing effect of the anchors. A marginally significant Anchor x Source interaction was also found, $F(1,72) = 3.62$, $p = .061$, $\eta_p^2 = .05$, showing that the anchoring effect was lower when the anchors were provided with a social source (i.e., other participant's answers) then when the same anchors were presented without a specified source.

Anchoring score. In order to create an anchoring score for each participant, trials that used low anchor values were reverse-scored. Participants' answers across both types of anchors were then averaged. Higher values in this composite score indicate stronger anchoring effects. Figure 1 shows the mean anchoring values across the experimental conditions. A 2 x 2 ANOVA with forewarning and source as between-subjects factors and anchoring score as dependent variable revealed a main effect of source, such that other-provided anchors led to lower anchoring scores (M = 0.47, 95% CI = [0.37, 0.56]) than the same anchors provided without a specified source (M = 0.61, 95% CI = [0.51, 0.71]), $F(1,73) = 4.36$, $p = .040$, $\eta_p^2 = .06$). A marginally significant main effect of forewarning was also found, $F(1,73) = 3.57$, $p = .063$, $\eta_p^2 = .05$, suggesting that forewarning participants about the potential biasing effect of the anchors led to less anchored answers ($M_{Forewarning}$ = 0.47, 95% CI $_{\text{Forewarning}}$ = [0.37, 0.57] vs. $M_{NoForewarning}$ = 0.61, 95% CI $_{\text{NoForewarning}}$ = [0.51, 0.70]). There was no Forewarning x Source interaction, F<1.

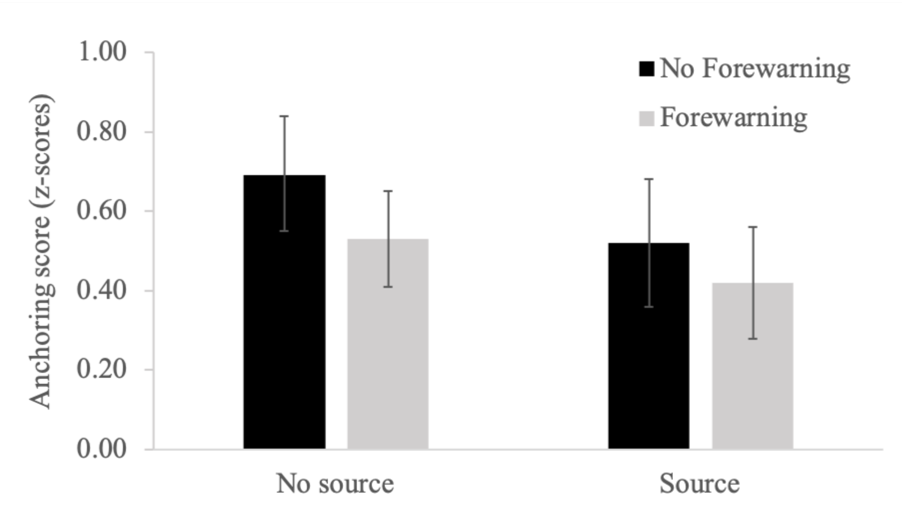


Figure 1. Mean anchoring values (in z-scores) as a function of the source of the anchor provided and the forewarnings about the anchoring effect. Note. Positive z-scores indicate responses less anchored answers. Error bars represent 95% confidence intervals of the mean.

The same set of planned orthogonal comparisons was carried out to test the hypothesis that the decrease in the anchoring effect should only occur when forewarnings were combined with socially provided anchors. The first planned comparison contrasting FS with all other conditions was significant, $t(73) = 2.07$, $p = .042$, $d = 0.55$, indicating that the combined effect of forewarning and presenting anchor values as other participants' estimates (FS) reduced the anchoring effect. However, the comparison contrasting NFS and FNS with NFNS also showed a significant decrease in the anchoring effect, $t(73) = 2.06$, $p = .044$, $d = 0.51$, suggesting that, at least under some circumstances, forewarning and source by themselves (and not only their combination as we expected) may reduce the anchoring bias. Finally, the comparison contrasting NFS with FNS was not significant, $t < 1$.

Response times. Figure 2 shows participants' mean response times (in milliseconds) as a function of the source of the anchor provided and the forewarnings about the anchoring effect. A 2 x 2 ANOVA with forewarning and source as between-subjects factors and mean response times as the dependent measure showed that participants in the forewarning conditions took longer to give their answers ($M = 5505.28$, 95% CI = [4885.77, 6124.79]) than those in the no forewarning conditions ($M = 4459.66$, 95% CI = [3865.79, 5053.52]), $F(1, 73) = 5.90$, $p = .018$, $\eta_p^2 = .08$.

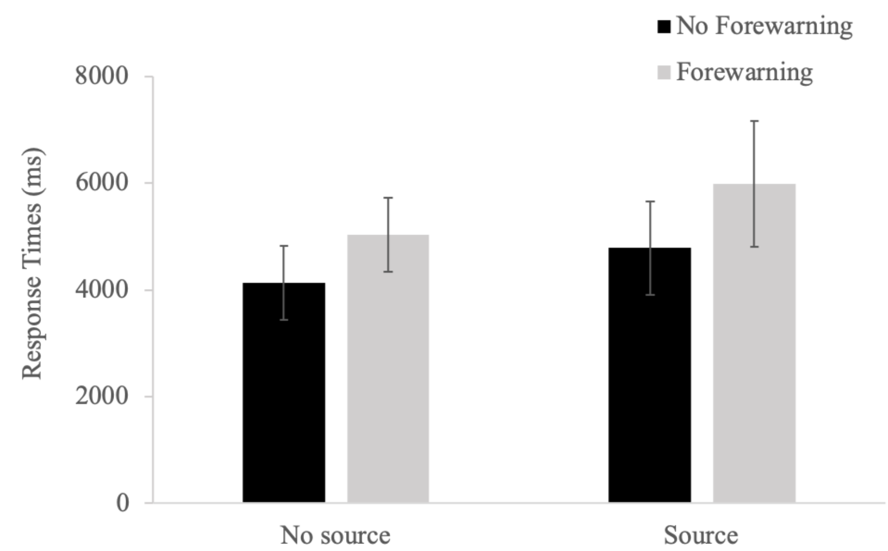


Figure 2. Mean Response times (in milliseconds) as a function of the source of the anchor provided and the forewarnings about the anchoring effect. Error bars represent 95% confidence intervals of the mean.

A marginally significant main effect of source was also found, $F(1, 73) = 3.45$, $p = .067$, $\eta_p^2 = .05$, suggesting that participants took longer to give their answers when the anchors were presented as other participant's answers (*M* = 5382.53, 95% CI = [4788.66, 5976.39]) than when the same anchors were presented without a specified source (*M* = 4582.41, 95% CI = [3962.90, 5201.92]). There was no Forewarning x Source interaction, $F<1$.

More relevant to the hypothesis concerning the combined effect of forewarning and source, planned orthogonal comparisons only revealed significant differences for the comparison contrasting FS with all remaining conditions, $t(73) = 2.67$, $p = .010$, $d = 0.71$. No significant differences were found for the comparison contrasting NFS and FNS with NFNW, $t(73) = 1.47$, $p = .147$, $d = 0.36$, nor for the comparison contrasting NFS with FNS, $t < 1$.

Furthermore, the correlation between response times and anchoring was negative and highly significant, $r(78) = -.36$, p=.001. That is, as response times increased response bias towards the anchors decreased.

Supplementary analyses. For the sake of brevity, analyses of the metacognitive measures are not reported here since they do not relate directly to the hypotheses under test. The interested reader can find these exploratory analyses in Appendix B1.4.

Discussion

Presenting the anchors as answers given by other peers (source conditions) substantially reduced the anchoring effect. Informing participants about the anchoring effect (forewarning conditions) marginally reduced anchoring. Planned comparisons to test for the combined effect of Forewarning and Source showed that the reduction of the anchoring effect was particularly striking when participants were previously forewarned about the anchoring effects and then presented with others' estimates as anchors (FS condition). However, the residual planned comparison contrasting NFS and FNS with NFNS conditions also showed a significant decrease in the anchoring effect. In other words, anchoring reduction in this experiment was not fully accounted by the combined effect of social source and forewarning.

Response times were longer for the forewarning condition and marginally so for the source condition. However, planned comparisons qualified these results by showing that the combined effect of source and forewarning (and none of the remaining comparisons) increased response times. In other words, deliberate and effortful adjustment was predicted and found to occur as the result of treating others as biased sources of information that need to be corrected.

Given that the effectiveness of NFS and FNS in reducing anchoring was not predicted, we ran a replication of this first experiment with enough power to check for the robustness of the effects here found.

Experiment 2

This experiment aimed at replicating and generalizing the results of Experiment 1 to a new sample, using a different set of general knowledge questions (see Appendix B2.1.). Small differences were introduced in the experimental procedure to clarify to what extent they can account for the unexpected success of NFS and FNS conditions in reducing the anchoring effect. Specifically, Experiment 2 uses a clear forewarning of the anchoring effect (without including an illustration of it); and the source of the anchors in the Source Conditions was more explicitly identified (see Appendices B2.2. and B2.3.).

Method

Participants. The sample size was determined a priori to replicate the combined effect of Forewarning and Source obtained in Experiment 1 ($d = 0.55$). To detect an effect of the same magnitude in a one-tailed t-test, with $\alpha = 0.05$, and power = 0.80, 108 participants were needed.

Given that data were collected online, we recruited some extra participants to deal with the eventual need to exclude participants from the sample. Thus, one-hundred twenty participants (64 Female; $M_{age} = 33.76$, $SD = 9.63$) were recruited via Amazon's Mechanical Turk (MTurk) online platform.

Materials. Participants were presented with 10 general knowledge questions taken from Jacowitz and Kahneman (1995). As in Experiment 1, each of these questions was accompanied by one possible response (the anchor value), which in half of the trials was a low value (low anchor), and in the other half was a high value (high anchor). These values corresponded to the 15th and 85th percentiles of Jacowitz and Kahneman's (1995) calibration group. The questions were presented in a randomized fixed order, and two different versions of the material were used to control for content effects. Knowledge questions presented with high anchors in version A were presented with low anchors in version B and vice-versa.

Procedure. The procedure was similar to that of Experiment 1, except that data collection was done online. The following forewarning instructions were provided to participants: "Previous research has demonstrated that when facing different kinds of questions we are often influenced by possible responses that are provided to us. The goal of this study is to evaluate to what extent you can avoid this tendency. Specifically, to what extent can you resist being influenced by external sources of information when questioned about general knowledge themes.". Also, the source of the anchor was more explicitly identified, while in Experiment 1 the source was subtly suggested in the question: "Do you agree with the above presented answer [vs. this participant's answer]?", in Experiment 2 the answers were presented with a label that explicitly identify them as provided by another participant: "James' answer was: …"; or without a specified source: "Answer: …". Additionally, each trial was composed by one single question. Specifically, in each trial participants were presented with a general knowledge question along with a possible (but not necessarily correct) answer that was either provided by a previous participant in the study (NFS and FS conditions) or without a specified source (NFNS and FNS conditions). Participants' task was to give their own answer to the presented question, indicating how confident they were about their answers (in 9 points rating scale from 1- *not at all to* 9 – *totally confident*) (see Appendices B2.2. and B2.3. for the instructions and an illustration of the trials).

After answering the 10 general knowledge questions, participants were asked to indicate: a) to what extent they believed to have been influenced by the responses provided to the general knowledge questions (from 1 – *not influenced at all*, to 4 – *totally influenced*) (subjective influence); b) how knowledgeable they were about the questions' subject matters (from 1- *none*, to 4 – *very much*) (subjective knowledge); and c) how they perceived their overall performance (from 1 – *very poor*, to 4 – *very good*) (subjective performance). As in Experiment 1, these measures, were included to explore how participants' metacognitive experiences relate with their objective performance in the anchoring task. Due to the online nature of data collection and programming issues, in was not possible to register response times in a systematic and reliable way, making it impossible to analyze response times in this experiment.

Results

As in Experiment 1, before analyzing participants' responses and calculating the correspondent anchoring scores, we excluded non-interpretable and nonsensical answers (e.g., answering that the telephone was invented in 88 or that the height of Mount Everest is 30 feet). Less than 5% (4.25%) of the responses were excluded using this criterion.

Participants' answers were then transformed into z-scores, separately for each question. Answers following low and high anchors were then averaged into two composite scores (low and high anchors estimates).

Low and high anchors. A 2 x 2 x 2 mixed ANOVA with anchor (low and high) as a within-subjects factor, and forewarning (with and without) and source (with and without) as between-subjects factors showed a large main effect of anchor, such that mean answers were lower for the low anchor trials (M = - 0.48, 95% CI = [-0.56, -0.40]) when compared to the high anchor trials (*M* = 0.50, 95% CI = [0.40, 0.60]), $F(1,116) = 223.26$, $p < .001$, $\eta_p^2 = .66$. This main effect was qualified by a marginally significant Anchor x Source interaction effect, $F(1, 116) = 3.87$, $p = .051$, $\eta_p^2 = .03$, such that the difference between high and low anchors (i.e., the anchoring effect) is larger for the no source than for the source conditions. No other significant effects were found.

Anchoring score. Participants' answers to low and high anchors trials were aggregated in one composite score by reverse-scoring the low anchor values (see Figure 3). Higher values in this score indicate stronger anchoring effects.

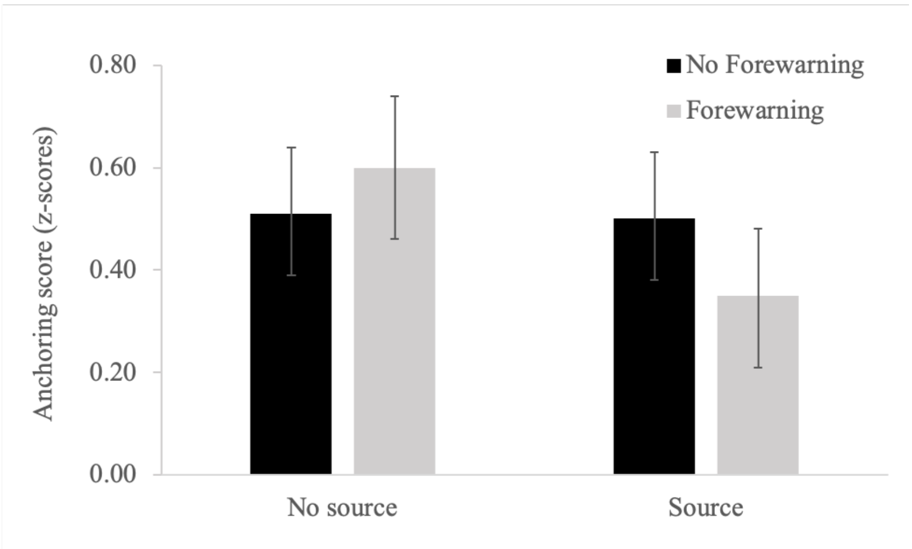


Figure 3. Mean anchoring values (in z-scores) as a function of the source of the anchor provided and the forewarnings about the anchoring effect. Positive z-scores indicate responses less anchored answers. Error bars represent 95% confidence intervals of the mean.

A 2 x 2 ANOVA with forewarning (with and without) and source (with and without) as between-subjects factors revealed a main effect of source, $F(1, 116) = 4.13$, $p = .045$, $\eta_p^2 = .03$, such that the anchoring score was lower when the anchors were said to be provided by another person (M = 0.43, 95% CI = [0.34, 0.52]) than when the same anchors were provided without a source (M = 0.56, 95% CI = [0.47, 0.65]). This effect was qualified by a marginally significant Source x Forewarning interaction, $F(1, 116) = 3.50$, $p = .064$, $\eta_p^2 = .03$. While in the no source conditions, forewarning participants about the anchoring effect slightly increased it, in the source conditions the same forewarning led to a significant decrease in anchoring. A marginally significant main effect of forewarning was also found, such that participants who were forewarned about the effect gave less anchored answers (M = 0.47, 95% CI = [0.33, 0.61]) than those who were not forewarned (M = 0.51, 95% CI = [0.38, 0.64]), $F(1, 116) = 3.50$, $p = .064$, $\eta_p^2 = .03$.

The same set of planned orthogonal contrasts used in Experiment 1 was carried out to test the hypothesis that the reduction of the anchoring effect should only occur when combining forewarning about the effect with specifying the source of the anchoring value as other participants' estimates (FS). As predicted, only the comparison that contrasted the FS condition with all other conditions showed significant differences, $t(116) = 2.58$, $p = .011$, $d = 0.55$. The two remaining comparisons were not significant - FNS and NFS vs. NFNS, $t < 1$; FNS vs. NFS, $t(116) = 1.04$, $p = .299$, $d = 0.48$.

Confidence ratings. Participants reported relatively low confidence levels across all conditions, with mean values below the scale mid-point (all ts = [3.02; 4.62], ps < .05). A 2 x 2 ANOVA with forewarning (with and without) and source (with and without) as between-subjects factors revealed no significant effects, all Fs = [0.11; 1.12], ps = [.292; .736] (see detailed results in Table 3 in Appendix B2.4.).

Participants' confidence was positively correlated with their subjective knowledge about the topics covered by the question, $r(120) = .73$, $p < .001$, suggesting that those who consider themselves more knowledgeable were more confident about their performance (or vice-versa). However, participants' performance (i.e., anchoring effects) was not significantly correlated with confidence (see Table 4 in Appendix B2.4.).

Discussion

This experiment successfully replicated and generalized Experiment 1 results to a new sample of participants, using a different set of general knowledge questions and a slightly different procedure.

The reduction of the anchoring effect was achieved when participants were provided with anchors from a social source (NFS and FS conditions) compared to the same anchors provided without a specified source. However, in support of our hypothesis, planned contrasts showed that this was only the case when participants were forewarned about the anchoring effect (FS condition). Contrarily to Experiment 1 results, neither forewarning participants about the anchoring effect without identifying the source of the anchor (FNS) nor presenting an answer from a social source without forewarning participants about the effect (NFS) were sufficient to reduce anchoring, which is consistent with most previous research (e.g., Epley & Gilovich, 2005).

General Discussion

Building on insights from research on people's naïve theories of bias (Ehrlinger et al., 2005; Pronin, 2007; Pronin et al., 2002, 2004; Wilson & Brekke, 1994; Wilson et al., 1996), we proposed that if externally provided anchors are presented as other people's estimates, then participants will perceive these estimates as biased and adjust away from them, thus reducing anchoring effects. Given that people more easily acknowledge others' susceptibility to biases than their own (e.g., Pronin et al., 2002), we further predicted that the reduction of anchoring effects should be stronger when participants are forewarned about the potentially biasing effect of anchors in judgment before being presented with other people's estimates. Furthermore, since adjusting away from other people's estimates is a deliberate corrective process that takes time and effort, participants who are forewarned about how people often fall prey of anchoring bias, and then receive answers from other people, were expected to take longer to give their absolute answers.

To test for these hypotheses, participants were forewarned (or not) about the anchoring effect and responded to several general knowledge questions after considering possible answers allegedly given by a previous participant in the study or presented without a specified source.

Two experiments used different sets of general knowledge questions, different variations of the standard anchoring paradigm and different collection procedures (Experiment 1 tested participants in laboratory settings whereas Experiment 2 tested participants via an online platform). Results consistently showed across both experiments that providing anchors as others' estimates reduced anchoring. Further analyses using planned comparisons indicated, as predicted, that such reduction in anchoring was due to the combined effect of social source and forewarning. In other words, forewarning participants about the anchoring effect reduced the bias when the anchors were provided as other participants' estimates. Furthermore, the response-time analysis in Experiment 1 (planned comparisons) suggests that such attenuation of anchoring was the result of effortful adjusting.

It is worth noting that planned comparisons in Experiment 1 (but not in Experiment 2) also showed evidence of bias reduction when contrasting both the forewarning and no source condition (i.e., participants were forewarned about the anchoring effect and anchors were provided without a specified source) and the no forewarning and source condition (i.e., no forewarning and anchors presented as other participants' estimates) with the no forewarning and no source/control (i.e., no forewarning and anchors presented without a specified source). However, this result did not replicate in Experiment 2. Only forewarnings about anchoring combined with anchors presented as other people's estimates reliably reduced the anchoring effect across both experiments. Future research should further explore the boundary conditions for a successful reduction of the anchoring effect.

Overall, the reported results are aligned with more recent research on anchoring (Blankenship, Wegener, Petty, Detweiler-Bedell, & Macy, 2008; Simmons, LeBoeuf, & Nelson, 2010; Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010; Wegener, Petty, Detweiler-Bedell, & Jarvis, 2001). According to Simmons et al.'s integrative theory of anchoring, the reason why people do not often adjust from externally provided anchors is that they do not know how to make the adjustment. Self-generated and externally provided anchors differ in the extent to which they cue the correction/adjustment processes. Self-generated anchors are known to be close to the correct answer but in need of an upward or downward adjustment. For instance, to answer the question "what is the freezing point of vodka?" one may think about the freezing point of water, and then adjust downwards. In contrast, externally provided anchors have been described in most previous research as randomly generated numbers presented without a specified source, giving participants no clue about its potential biasing effect on their answers. In agreement with this rational, Simmons et al. (2010) showed that incentives for accuracy successfully reduced anchoring with both self-generated or externally provided anchors, provided that participants are provided with cues on how to adjust from the anchors.

According to the attitudinal approach to anchoring (Blankenship et al., 2008; Wegener et al., 2001; Wegener et al., 2010), the standard anchoring paradigm is analogous to other persuasion contexts, in that participants are exposed to a piece of information (the anchor) and explicitly asked to take it into consideration. As it happens with other persuasion factors (e.g., source characteristics), anchors may serve multiple roles: sometimes they work as simple cues that directly influence people's estimates; other times they may influence effortful processing. When people have some reason to disagree with the provided anchor, a search for disconfirmatory information will be triggered and the influence of the anchor will be lessened. For instance, when the anchors come from a less credible/competent other, one should expect less anchoring than when the same anchors come from a more credible/competent other. Indeed, the credibility of the source has been shown to influence the extent of its influence. Participants' numeric estimates were more anchored in high credibility conditions than in the low credibility ones (Dowd et al., 2014; but see Langeborg & Eriksson, 2016).

Our own perspective stems from research on naïve theories of biases showing that although people recognize the existence and the impact of most cognitive and social biases in others' judgments, they often lack the recognition of these same biases in their own judgments (Ehrlinger et al., 2005; Pronin, 2007; Pronin et al., 2002; 2004). Based on this view, we predicted and found that presenting anchor values as estimates of other persons leads participants to adjust away from these others' estimates particularly when participants are forewarned of the potentially biasing effect of anchors in judgment. Our perspective is also aligned with Mercier and Sperber (2011, 2017; Mercier, 2014; Mercier & Landemore, 2012; Sperber et al., 2010) suggestion that the main function of reasoning is to devise and evaluate arguments intended to persuade, and that people exert epistemic vigilance towards others' judgments as a way for avoiding misinformation (Sperber et al., 2010). In this vein, forewarning about anchoring effects may be seen as a way to make people more sensitive to the possibility of being misinformed by anchor values when these are suggested by others.

Regardless of the different (and partially overlapping) aforementioned theoretical approaches, our results contribute to extant research by exploring new circumstances under which externally provided answers will trigger adjustment (high-elaboration) processes that were initially associated with self-generated anchors (Epley & Gilovich, 2001, 2005). Specifically, since people tend to perceive others as more prone to bias than themselves, they adjust their responses away from these others' estimates. It thus seems that the so-called bias blind spot (Pronin et al., 2002) may help attenuate anchoring under appropriate circumstances.

In addition, our results suggest the advantage of studying the combined effects of heuristics rather than exploring them in isolation. In the present case, it is the bias blind spot that prevents participants from using diagnostic information (i.e., forewarnings) to correct the effects of the anchoring heuristic. In other words, it is not so much that anchoring is immune to forewarning (as sometimes suggested in the literature; e.g., Wilson et al., 1996), but rather the belief that others (but not the self) are prone to biases (i.e., the bias blind spot) that renders forewarning useless. Indeed, it is likely that in many real world settings, judgment biases result from the interaction of several heuristics and not from their isolated operation (for similar exemplars on the combined effect of heuristics regarding confirmation biases see Klayman, 1995; Nickerson, 1998).

Finally, recent research has also compared anchoring and advice-taking paradigms (Fiedler, Hütter, Schott, & Kutzner, 2019; Rader, Soll, & Larrick, 2015; Schultze, Mojzisch, & Schulz-Hardt, 2017). In both anchoring and advice-taking tasks, participants are asked to make estimates after being exposed to other numerical values (i.e., anchors or advices), with the dependent variable being the extent to which participants' estimates are influenced by these values. Whereas the assimilation of one's judgment to the anchor is the classic outcome of anchoring research, people typically underestimate advices in advice-taking research. A crucial difference between these two paradigms is that the numerical values presented in the advice-taking task are often embedded in a communicative context in which participants attempt to learn something from the advice giver (Fiedler et al., 2019), while the anchors used in the anchoring task are often presented in a social vacuum as randomly generated numbers without a specified source.

By using socially-provided anchors, our experiments come closer to advice-taking tasks. However, rather than promoting conditions for the use of advice coming from an advice-giver, our manipulations aimed at exploring to what extent participants could avoid the influence of "others' estimates" presented as a more causal source of information instead of advices coming from advice-givers. As aforementioned, most previous research has shown people's tendency to underweight useful advice (e.g., Soll & Larrick, 2009; Yaniv, 2004; Yaniv & Kleinberger, 2000), but less is known about how people treat advice perceived as useless or biased (Fiedler et al., 2019). Future research using experimental paradigms similar to the ones used here (i.e., embedding anchoring effects in social contexts) could contribute to better understanding how others' estimates influence one's judgments depending on being perceived as just a possible response or an advice (see Schultze et al., 2017).

Chapter 4 – Reasoning About Others' Heuristic and Rule-Based Answers

When faced with judgment tasks, people use heuristics that are qualitatively distinct from the prescriptions of normative models (Kahneman et al., 1982; Tversky & Kahneman, 1974). Although very efficient under many circumstances, these heuristics have been shown to lead to a variety of systematic biases (Einhorn & Hogarth, 1981; Fischhoff et al., 1977; Gilovich et al., 2002; Kahneman et al., 1982; Keren & Teigen, 2004; Koehler & Harvey, 2008; Nisbett & Ross, 1980; Slovic et al., 1977). For instance, people often evaluate the quality of arguments based on their prior beliefs instead of following logical rules (e.g., Evans et al., 1994; Evans et al., 1983; Klauer et al., 2000; Markovits & Nantel, 1989; Thompson & Evans, 2012) they fail to reason disjunctively (e.g., Shafir, 1994; Toplak & Stanovich, 2002) and tend to maximize relative savings at the expense of absolute savings (e.g., Bartels, 2006; Mata, 2016).

However, most of the previous research on heuristics and biases has treated human judgment as an individual process, neglecting the fact that many of our decisions take place in social settings, where people influence and are influenced by others' responses and reasoning

(Larrick, 2016). In other words, by focusing mainly on people’s responses to socially decontextualized reasoning tasks, researchers have payed less attention to the interaction between the cognitive processes underlying judgment biases and the social contexts where people´s reasoning and responding unfolds (Butera, Legrenzi, & Oswald, 1997; Hogarth, 1981; Mercier & Sperber, 2011).

Our goal is to contribute to close this research gap by testing how people’s performance in classic reasoning tasks is affected when these tasks are presented in a social context in which people can use others’ answers to guide their own reasoning.

Reasoning in Social Contexts

A distinctive feature of the reasoning tasks typically used to study human judgment (in the tradition of heuristics and biases research program) is that they trigger compelling intuitive answers (i.e., heuristic-based answers) that are in conflict with the rule-based (and often correct) ones. To appropriately answer these tasks, people need to override the heuristic answers by engaging in deliberate reasoning (for a review, see Evans, 2007). Several variables have been found to influence the likelihood of engaging in deliberate reasoning: reasoners’ cognitive abilities (e.g., De Neys, 2006; Stanovich, 1999), the amount of time allocated to the task (e.g., Evans & Curtis-Holmes, 2005; Finucane et al., 2000), the instructions provided (Daniel & Klaczynski, 2006; Evans et al., 1994; Vadeboncoeur & Markovits, 1999), and metacognitive cues (e.g., answer and perceptual fluency; Alter et al., 2013; Alter et al., 2007; Thompson et al., 2011; Thompson et al., 2013). We propose that the social contexts where reasoning takes place may also affect people’s engagement in deliberate reasoning.

People often depend on the information provided by others when making decisions in social contexts, which leaves them vulnerable to the risk of being accidentally or intentionally deceived by others. Being able to identify who should be trusted and who should be treated with suspicion is thus crucial to avoid relying on inappropriate information (Kruglanski et al., 2005; Mercier & Sperber, 2011; Sperber et al., 2010).

According to the epistemic vigilance approach, humans have a suit of cognitive mechanisms, targeted at the risk of being misinformed by others (Mercier & Sperber, 2011, 2017; Sperber et al., 2010). People are vigilant towards the information contents (i.e., is the information provided coherent with previous beliefs?) and towards who provides the information (i.e., is the social source competent and/or benevolent?). When the communicated information in not coherent with previous beliefs, people either reject the information or revise their previous beliefs. Since people calibrate their trust on the information provided by others based on how competent (and benevolent) others are perceived (see also Fricker, 1995; Mayo, 2015; Mayo, Alfasi, & Schwarz, 2014; Petty & Cacioppo, 1984, 1986), the extent to which people accept or reject this information strongly depends on how much they trust its source. It follows that information incoherent with prior beliefs is more easily accepted when it comes from a highly trusted source. On the contrary, information that is coherent with people's prior beliefs goes under the radar of epistemic vigilance mechanisms, being accepted independently of its source.

Building on this research, Mata, Fiedler et al. (2013) suggested and found that people's disposition to be vigilant towards the information provided by others can help them to second guess others' heuristic responses, and consequently to more likely reject these responses and come up with logical or statistically correct responses. In a set of studies participants were asked to solve several classic reasoning problems that present a conflict between a highly appealing (but wrong) heuristic answer and a (correct) rule-based answer that requires the use of very basic logical or statistical principles. These problems were presented together with heuristic-based answers allegedly given by previous participants in the study (or without specifying the source). Participants' tendency to believe that others are more prone to biases than themselves (the bias blind spot, BBS; Pronin, 2007; Pronin, 2002, 2004) was used as a measure of their disposition to be vigilant towards others' responses. Results from these studies showed that participants who perceived others as more biased than themselves (i.e., showed the BBS) were better at a) rejecting heuristic answers; and b) giving the logical/statistically correct response, when these heuristic answers came from other participants, but not when the same answers were provided without a social source. In contrast, participants who believed that others were less prone to biases than themselves, more often failed to detect flawed heuristic responses. As a consequence, they tended to adopt others' heuristic answers more often than when the same answers were provided without a social source.

While this previous research focused on individual differences in BBS as estimates of the disposition to be vigilant towards others' reasoning, in the current paper we are interested in testing whether the likelihood of adopting a critical mindset when evaluating others' reasoning can also be situationally induced. Specifically, we aim to test whether people are more vigilant towards reasoning mistakes when these mistakes (i.e., heuristic answers) are given by others perceived to be low in competence, compared to when the same mistakes are given by others high in competence.

Overview of the experiments

In two experiments participants are presented with several reasoning problems. Before responding to each problem, they receive the answer allegedly given by another participant in the study. Depending on the experimental condition, the alleged participant is suggested to be high (vs. low) in competence through the presentation of an individual profile. The provided answers are either the heuristic but incorrect answers that most people give to these reasoning problems (Experiment 1) or a mixed set of heuristic and rule-based answers (Experiment 2). Both experiments included a second set of problems presented without answers from other participants to test whether the cognitive effects of reasoning about others' reasoning spill over to subsequent reasoning problems.

If participants are more critical towards the answers provided by a low competent other, then they will more likely detect the incorrect solutions and perform better when answering in the low competence than in the high competence condition (Hypothesis 1). As a consequence of critically rejecting other participant's answers in the first set of problems, participants in the low competence condition should also have a better performance in the second phase (when no answers from others are provided) (Hypothesis 2).

Experiment 1

Method

Participants. One-hundred forty-eight participants (75 females, M_{age} = 33.22, SD = 10.70) were recruited using Prolific Academic online platform. The study lasted for approximately 25 minutes, and participants received £2.15 for their participation. All participants were native English speakers with at least 18 years old. Six participants reported previous knowledge about the materials and/or suspicion of the manipulation and were therefore excluded from the analyses. The final sample included 142 participants. Prior approval for running the experiments reported in this paper was obtained from the Ethics Committee of the Faculty of Psychology, University of Lisbon.

Materials.

Reasoning problems. Participants' ability to detect and to overcome biases in others' reasoning was measured using five different types of reasoning tasks (i.e., syllogisms, transitive reasoning problems, semantic illusions, proportion dominance problems, and disjunctive insight problems). All of these tasks have in common the fact that they trigger a compelling intuitive and incorrect (heuristic-based) answer that is at odds with the correct, rule-based answer[8]. To correctly answer these tasks, participants need to inhibit the heuristic answer and replace it for the rule-based one. Moreover, according to Stanovich's taxonomy of rational thinking errors in heuristics and biases tasks, all these tasks are more process dependent than knowledge dependent (Stanovich, 2016; Stanovich, Toplak, & West, 1998). The fact that they do not require sophisticated knowledge of formal decision rules from the decision-maker, makes these tasks suitable to test for improvements in judgmental accuracy resultant from our experimental manipulations (see Arkes, 1991; Lerner & Tetlock, 1999; Simonson & Nye, 1992).

The reasoning problems were evenly distributed across five blocks of trials in a randomized fixed order. The first four blocks of trials corresponded to the first phase of the experiment, in which participants were allegedly paired with other participant in the study, while the last block corresponded to the second phase, in which participants were answering on their own (i.e., without access to other's answers).

In the first phase of the experiment, each reasoning problem was presented with an answer that was allegedly given by another participant in the study. Sixteen of the twenty reasoning problems in this phase were presented in their conflict version (i.e., heuristic and rule-based processes produce conflicting outputs). In all these trials, the answer provided by the other participant was the heuristic and incorrect one. In the remaining four trials (fillers), the reasoning problems were presented in their no-conflict version (i.e., heuristic and rule-based processes converge in the same answer) and thus the other participant's heuristic answers for these problems were correct because they coincide with the rule-based answers. These

[8] This is the classical format used in the HB tasks and corresponds to what became known as the 'conflict version' of these tasks. In the 'no conflict version' the heuristic answer coincides with the rule-based one.

fillers (i.e., no-conflict versions) were included to avoid that the correct answer would always imply rejecting other participant's answer. The five problems included in the second phase of the experiment were presented in their conflict version and without an answer from a previous participant. Two versions of material were used to control for order effects (see Appendix C1.1. for the reasoning problems used in this experiment).

Syllogistic reasoning problems. The syllogisms used were adapted from De Neys and Franssens (2009). Following the paradigm introduced by Evans et al. (1983), the syllogisms presented to participants consisted of two premises and a stated conclusion. The syllogisms were presented either in their logically valid (affirming the antecedent - Modus Ponens) or invalid (affirming the consequent) forms, and with conclusions that were either believable or unbelievable. Three of the conflict-version syllogisms presented valid arguments with unbelievable conclusions, and one presented invalid arguments with a believable conclusion. In the no-conflict version, the syllogisms presented valid arguments with a believable conclusion. Participants' task was to evaluate the validity of the conclusion, and the instructions emphasized that this should be based on logical reasoning rather than on prior beliefs. The heuristic answer consists on accepting as valid a believable but logically invalid conclusion or on rejecting as invalid an unbelievable but logically valid conclusion.

Transitive reasoning problems. Following Banks and Hope (2014), the transitive reasoning problems used consisted of three premises and a conclusion. As with the syllogisms, the logical validity (valid/invalid) and believability (believable/unbelievable) of the conclusion were manipulated. The logical validity of the conclusion was manipulated by changing the premises so that the same conclusion was either valid or invalid. A nonsense term was included in the premises to avoid that they conflicted with participants' background knowledge. The believability of the conclusion was manipulated by reversing the elements in the conclusion. Two of the conflict-version problems presented valid arguments with unbelievable conclusions, and two others presented invalid arguments with believable conclusions. In the no-conflict version, the arguments were presented in a logically valid form with a believable conclusion. Participants' task was to indicate whether the conclusion followed logically from the premises, assuming the premises were true. The heuristic answer consists on accepting as valid a believable but logically invalid conclusion or rejecting as invalid an unbelievable but logically valid conclusion.

Semantic illusions[9]. The illusory sentences used were taken from Erickson and Mattson (1981), Mata, Ferreira, and Reis (2013), and van Oostendorp and De Mul (1990). These sentences were created by replacing a true term in the original sentence with a semantically similar but incorrect term. To illustrate, the sentence "The sheep is a farm animal that provides milk, cheese and wool", that is true, becomes illusory when the word "sheep" is replaced by "goat": "The goat is a farm animal that provides milk, cheese and wool". Four sentences were presented in their illusory version (conflict-version) and one in its true version (no-conflict). Participants' task was to evaluate whether sentence is true or false. The heuristic answer on this task consists on accepting the illusory sentences as true.

Proportion dominance. Proportion dominance problems were taken from Bartels (2006) and Mata (2016). In these problems, participants were presented with different scenarios involving decisions to save the lives of people or animals. In each scenario participants were presented with two options that forced a tradeoff between absolute and relative savings (i.e., saving more lives vs. saving a larger proportion of a population), and their task was to indicate which of the two options they considered the best. In four of these problems (conflict version), the option that involved saving more lives is the one that saves a smaller proportion of the population (e.g., saving 225 lives in 230 = 75% vs. 230 in 920 = 25%). In the

[9] Although semantic illusions are not reasoning problems, they also prompt a highly intuitive answer that is contrary to the correct one.

no-conflict version, the option that saves more lives is also the one that saves a larger proportion of the population. The heuristic answer in these problems corresponds to a preference to maximize relative savings at the expense of absolute savings (i.e., saving 225 lives in 230 = 75% rather than 230 in 920 = 25%, in the abovementioned example).

Disjunctive insight problems. Disjunctive insight problems were based on Sequeira, Ferreira, and Almeida (2013), and Toplak and Stanovich (2002). In these problems participants were presented with sentences that explicitly describe a set of relations between subjects/objects, and were asked about a another relation that requires to take into consideration information that is not explicitly stated (e.g., "Jack is looking at Ann but Ann is looking at George. Jack is married but George is not. Is a married person looking at an unmarried person? a) Yes, b) No, c) Cannot be determined"). To derive the correct conclusion participants have to consider both possibilities for the non-stated relation. In the abovementioned example, participants need to consider both the possibility of Ann being married or unmarried: if Ann is married, then the answer is "Yes" because she would be looking at George who is unmarried; if Ann is not married, the answer is still "Yes" because Jack, who is married, would be looking at Ann. The heuristic answer in these problems consists of choosing the option "Cannot be determined". In the no-conflict version of these problems, participants were asked about a relation that was explicitly stated in the problem.

Cognitive Reflection Test. An adapted version of the Cognitive Reflection Test (CRT; Frederick, 2005) with different contents was used to control for prior exposure to the test, and to avoid that participants could look up the answers on the internet (see Goodman, Cryder, & Cheema, 2013). The CRT is frequently used as a measure of people's reflective thinking skills, with high performance in the test being associated with a high propensity to engage in more effortful processing and lower susceptibility to judgmental biases (Frederick, 2005; Oechssler, Roider, & Schmitz, 2009; Toplak, West, & Stanovich, 2011, 2013). The CRT was included in this experiment to explore the relationship between participants' cognitive reflection and their propensity to follow or reject other participant's answers. Higher tendency to reflect upon the information provided (i.e., higher performance in the CRT) is expected to be associated with less reliance on other participant's answers, independently of her competence.

Other participants' profiles. Two profiles were created to manipulate other participant's competence. In the high competence condition, the other participant was presented as a 35 year-old male, named David, who was a consultant in artificial intelligence and robotics. His hobbies were reading sci-fi books, building drones, programming, and watching TV series. This profile was presented with a picture of a man looking at a blackboard with mathematical formulas. In the low competence condition, the participant was presented as a 19 year-old female, named Sarah, who was a babysitter and whose hobbies were shopping at the mall, partying with friends and going to makeup and fashion events. Sarah's profile picture portrayed a female taking a fun selfie on the beach (see Appendix C1.2.).

Design. The experiment followed a design with 2 (other participant's competence: high vs. low) x 2 (cognitive reflection: low vs. high) conditions manipulated between participants.

Procedure. The experiment was set up on Qualtrics and launched on Prolific Academic Website as a study about how people answer reasoning problems in social contexts. Participants were randomly assigned to one of the experimental conditions (low vs. high competence) and received some general instructions about the experiment: at first, they would be paired with another participant from the study and then, working in turns, they would be either answering several reasoning problems or evaluating the other participant's answers to these same problems. The roles played by each participant in the first phase of the experiment (i.e., either respondent or evaluator) were said to be randomly chosen. In fact, in this first phase, participants were always instructed to evaluate other participants' answers, indicating whether they were correct or incorrect and then giving their own answer to each problem. Ratings of confidence (on a scale ranging from 1 – *not at all confident* to 9 – *totally confident*) were also requested for participants' own answers to each reasoning problem.

To make the manipulation credible, before starting the study participants were asked to fill in some demographic information (age, gender, occupation, hobbies) and to select a profile picture that would serve to introduce them to the other participant in the study. After completing this initial information, participants were (allegedly) paired with another participant in the study and received the correspondent background information (profile), which was manipulated to induce high vs. low competence (see Materials section). While waiting to be paired with other participant and before receiving the correspondent profile information, participants were asked to complete a modified version of the CRT.

The experiment consisted of two phases. In the first phase, participants were presented with 20 reasoning problems together with the heuristic answer given by the other participant for each problem. Participants' task was to indicate whether the other participant's answer was correct or incorrect and to provide their own answer to the same problem, stating how confident they were about their own answers (see Appendix C1.3. for the detailed instructions and an illustration of the trials).

At the beginning of the second phase, participants were told that they would be working alone in a new set of five reasoning problems. These problems were similar to those found in the first phase, but this time they were presented without an answer from the other participant. Participants' task was to answer each reasoning problem, stating how confident they were about their answers, using a scale ranging from 1 - *not at all confident* to 9 - *totally confident*. This second phase was included to test to what extent following or rejecting other participant's answers in the first phase would impact one's own reasoning.

After finishing the second phase, participants were asked to evaluate several occupational groups in the social and intellectual dimensions. We were interested in how participants evaluated AI & Robotics consultants' and babysitters' competence but other occupational groups (and the social dimension) were also included to make this goal less obvious (see Appendix C1.4.).

At the end of the experiment, participants were asked about their previous experience with the presented reasoning problems and were encouraged to leave their comments about the experiment. In the end of the experiment participants were fully debriefed and thanked.

Results

Manipulation check. A t-test for independent samples was used to test the effectiveness of our manipulation of source's competence. Results from this analysis showed that participants in the high competence (HC) condition perceived other participant's intellectual competence as higher ($M = 6.04$, $SD = 0.92$) than those in the low competence (LC) condition ($M = 4.38$, $SD = 0.96$), $t(140) = 10.47$, $p < .001$, $d = 1.77$.

Cognitive Reflection. Participants' performance on CRT (i.e., the number of items answered correctly) was taken as a measure of their tendency to engage in more elaborate reasoning rather than answer with a readily available intuitive response**.** This measure could vary between 0 and 3, with higher values revealing higher reflective reasoning. Participants' mean performance on CRT did not differ across the experimental conditions, $t < 1$, with both conditions exhibiting relatively low levels of performance ($M_{LC} = 1.20$, $SD_{LC} = 1.07$ vs. $M_{HC} = 1.26$, $SD_{LC} = 1.03$). A median split of the CRT performance (Median = 1.00) was used to divide the sample in two groups: Low vs. high CRT.

Accuracy. Figure 1 shows the proportion of correct answers for the first[10] and second phase problems as a function of other participant's competence, and participants' cognitive reflection.

[10] In the first phase, this proportion was calculated after excluding the four no-conflict trials.

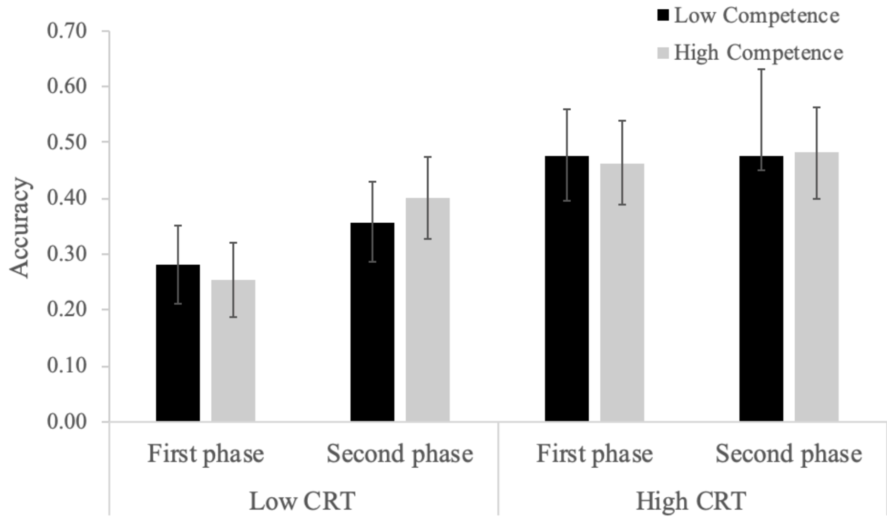


Figure 1. Proportion of correct answers as function of other participant's competence (low vs high) and participants' cognitive reflection (low vs. high CRT). Note: Error bars represent 95% confidence intervals of the mean.

The 2 x 2 x 2 mixed ANOVA with other participant's competence, and participants' cognitive reflection as between factors, and phase of experiment as a repeated measure revealed one main effect of participants' cognitive reflection, $F(1, 138) = 25.02$, $p < .001$, $\eta_p^2 = .15$, and one main effect of phase of the experiment, $F(1, 138) = 16.34$, $p < .001$, $\eta_p^2 = .11$. These two main effects were qualified by a marginally significant participants' Cognitive reflection x Phase of experiment interaction, $F(1, 138) = 3.51$, $p = .063$, $\eta_p^2 = .03$, such that low CRT participants slightly improved their accuracy from the first ($M = .27$, 95% CI = [.22, .31]) to the second ($M = .38$, 95% CI = [.33, .43]) phase, while high CRT participants maintained their accuracy levels in both phases ($M_{First} = .47$, 95% CI = [.41, .53] vs. $M_{Second} = .51$, 95% CI = [.45, .57]).

These results may be seen as an indication that participants with low CRT were more prone to uncritically following other participant's answers in the first phase problems, which lowered their performance in those problems. Once they stop receiving other's answers, in the second phase, their performance increased. Contrarily to our hypothesis, we found no main effect nor interactions with Other participant's competence, suggesting that this variable had no substantial impact on the extent to which participants followed or rejected other's answers.

We should note that the improvement in participants' accuracy between the two phases could also result from mere practice with the problems. To test for this alternative explanation, we splitted the first phase problems into two blocks (B1 and B2), and conducted a new 2 x 2 x 3 ANOVA with other participant's competence, and participants' cognitive reflection as between factors, and phase of experiment (B1 vs. B2 vs. Second phase) as a within factor. Assuming that practice effects increase performance in a progressive and linear fashion until reaching its ceiling, we should find significant differences between B1 and B2, in the first phase, and eventually (but not necessarily) between B2 and the second phase (i.e., when participants respond on their own without having access to other's responses). Results from this analysis revealed a main effect of participants' cognitive reflection, $F(1, 138) = 28.83$, $p < .001$, $\eta_p^2 = .17$, a main effect of phase of the experiment, $F(2, 276) = 15.04$, $p < .001$, $\eta_p^2 = .10$, and a significant participants' Cognitive reflection x Phase of experiment interaction, $F(2, 276) = 3.44$, $p = .034$, $\eta_p^2 = .02$.

More relevant for the testing of practice effects, comparisons between B1 and B2 showed significant differences for participants with high CRT ($M_{B1} = .43$, 95% CI = [.37, .49] vs. $M_{B2} = .51$, 95% CI = [.45, .57], $t(138) = 3.49$, $p = .001$, $d = 0.50$), but not for those with low CRT ($M_{B1} = .25$, 95% CI = [.20, .30] vs. $M_{B2} = .28$, 95% CI = [.23, .33], $t(138) = 1.33$, $p = .184$, $d = 0.15$). On the contrary, comparisons between B2 and the second phase revealed differences only for participants with low CRT ($M_{B2} = .28$, 95% CI = [.23, .33] vs. $M_{Second\ phase} = .38$, 95% CI = [.33, .43], $t(138) = 3.75$, $p < .001$, $d = 0.37$). No differences were found between these two blocks for participants with high CRT ($M_{B2} = .51$, 95% CI = [.45, .57] vs. $M_{Second\ phase} = .51$, 95% CI = [.45, .57], $t < 1$).

Taken together, these results are consistent with the existence of practice effects for high CRT participants but not so much for low CRT participants. They further suggest that uncritically following other's answers is likely to have contributed to the lower performance that low CRT participants showed in the first phase problems compared to the second one. Regardless, more research is needed to fully disentangle these two concurrent explanations: practice effects and uncritically following others' responses.

The correlation analyses revealed positive and significant correlations between participants' performance on CRT and their performance on both first ($r(142) = .48$, $p < .001$) and second ($r(142) = .33$, $p = .031$) phase problems. Participants' accuracy in the two phases of the experiment is also positively correlated, $r(140) = .56$, $p < .001$.

Confidence. Mean confidence was separately calculated for correct and incorrect answers. Figure 2 shows the mean confidence on correct and incorrect answers as a function of the other participant's competence, participants' cognitive reflection and phase of the experiment.

A 2 x 2 x 2 x 2 ANOVA with other participant's competence (low vs. high), and participants' cognitive reflection (high vs low CRT) as between factors, accuracy (correct vs. incorrect), and phase of the experiment as within factors, and confidence as dependent measure, revealed a main effect of phase of experiment, such that confidence was higher in the second ($M = 7.56$, 95% CI = [7.26, 7.86]) than in the first ($M = 7.42$, 95% CI = [7.14, 7.69]) phase problems, $F(1, 113) = 4.10$, $p = .045$, $\eta_p^2 = .04$; and main effect of accuracy, such that mean confidence was lower for correct ($M = 7.35$, 95% CI = [7.03, 7.67]) than for incorrect ($M = 7.63$, 95% CI = [7.36, 7.89]) answers, $F(1, 113) = 10.69$, $p = .001$, $\eta_p^2 = .09$. No other main effects or interactions were found.

Given that a correct answer in the first phase problems requires that participants dissent from other participant's answers, differences in confidence on correct and incorrect answers should be mostly found when participants were dissenting from heuristic responses given by a highly competent other but not when the same responses came from a low competence other. This should be particularly the case for low CRT (who we expected to be more susceptible to the influence of other participant's answers) but not for high CRT participants.

In fact, although other participant's competence did not interact with accuracy, low CRT participants in the high competence condition revealed lower confidence on correct (M = 7.05, 95% CI = [6.56, 7.54]) than on incorrect answers (M = 7.55, 95% CI = [7.17, 7.92]), $t(113) = 2.37$, $p = .019$, $d = 0.43$. For those answering in the low competence condition, no differences were found between confidence on correct (M = 7.33, 95% CI = [6.87, 7.79]) and incorrect answers (M = 7.47, 95% CI = [7.12, 7.82]), $t < 1$. Participants with high CRT showed no differences in their confidence levels on correct or incorrect answers, neither when answering in the HC nor in the LC conditions, all ts < 1.

This same pattern of results was found in the second phase problems, when participants were no longer presented with other participant's answers (see the bottom graph in Figure 2). These results suggest that the decrease in confidence experienced after rejecting answers from a highly competent other in the first phase, spillover to the second set of problems. Although problems in this second set were presented without other's answers, participants in the HC condition remained less confident on their correct than incorrect answers.

The correlation analyses revealed no significant correlations between participants' levels of cognitive reflection and their confidence on correct and incorrect answers.

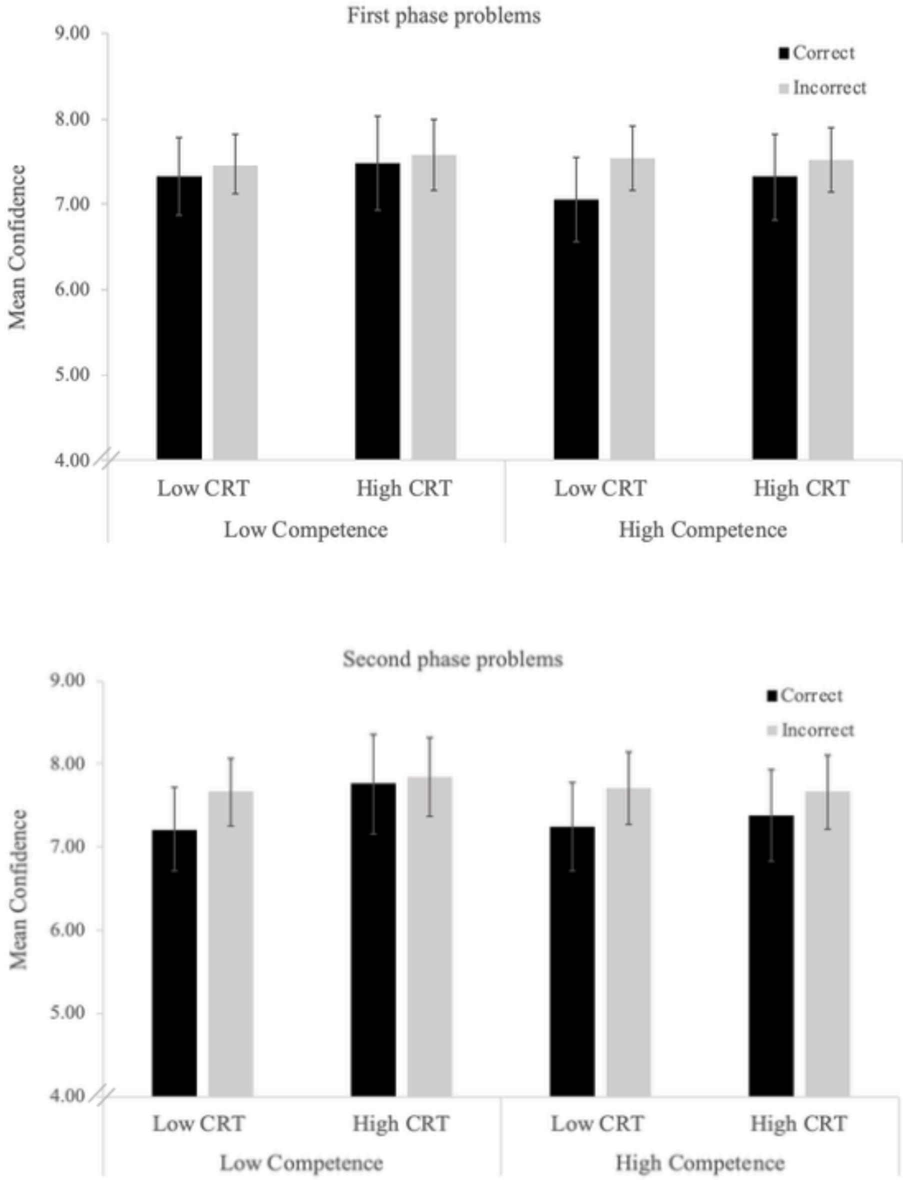


Figure 2. Mean confidence on correct and incorrect answers as a function the other participant's competence (low vs high), participants' cognitive reflection (low vs high CRT) and phase of the experiment (first phase problems on the top, second phase problems on the bottom). Note. Error bars represent 95% confidence intervals of the mean.

Discussion

Contrary to the hypothesis that people more likely second guess and overcome biased responses coming from others who are low (vs. high) in competence, results from this first experiment showed that participants' accuracy on the reasoning problems did not differ across the two other participant's competence conditions, regardless of participants' cognitive reflection. Comparisons of participants' performance between the two phases suggest that low CRT participants were mostly following other's answers in the first phase. Indeed, when they had to respond on their own to a second set of similar problems, their performance increased. No differences were found between the two phases for high CRT participants.

Furthermore, when the low CRT participants dissented from the high competence other's responses (i.e., when they rejected the heuristic response and gave the rule-based response) their confidence on their own (correct) responses decreased. Dissenting from the low competence other's response had no impact on participants' confidence levels. In other words, the competent source undermined participants' confidence in their own rule-based responses and this had carry-over effects to a second set of problems where participants responded on their own. The confidence levels of high CRT participants were not affected neither by the accuracy or their answers nor by the other participant's competence.

Several aspects of our experimental manipulations might have contributed to these results. First, our manipulation of other participant's competence did not completely succeed in clearly distinguish between two groups perceived either as high or as low in competence. Indeed, only half of participants in the LC condition considered babysitters/the group as low in competence[11].

[11] We conducted additional analyses excluding participants who evaluated the low competence group as high in competence and vice-versa. Results from these analyses replicate those found without participants' exclusion. We should note, however, that the two groups became unbalanced after participants' exclusion, which compromise the interpretation of these additional analyses.

Second, the fact that other participant's answers were always the heuristic ones could also have contributed to weaken our manipulation of competence of the source. Heuristic answers are the answers that most people give to these problems. If people's vigilance towards other's answers is lowered when these answers do not conflict with their own (coherence checking mechanism of epistemic vigilance), it is reasonable to assume that participants' acceptance or rejection of other participant's heuristic answers will not be affected by her competence, at least when there are no other reasons to be vigilant. This post hoc account of the results led us to run a second study where other's responses where sometimes the rule-based answers and hence incoherent with the dominant heuristic-based answers participants usually give to the kind of reasoning problems used here. This should be enough to trigger coherence checking mechanisms of epistemic vigilance and, as a result, lead participants to pay more attention to the source of the provided answers. Responses coming from a competent other should be followed whereas responses coming from an incompetent other should be second-guessed, more often rejected, and replaced by rule-based (correct) answers.

Experiment 2

This experiment aimed to continue testing whether people adopt a more critical mindset when evaluating the reasoning of less (vs. more) competent others, addressing the limitations of Experiment 1.

As aforementioned, according to the coherence checking mechanism of epistemic vigilance, the tendency to scrutinize others' answers might be lowered when these answers are not in conflict with one's previous intuitions. Such decrease in epistemic vigilance is likely to have happen for most participants in Experiment 1 because the other participant always gave the heuristic answers, which happen to be the default answers to the type of reasoning problems used. In contrast, if the provided answer contradicts one's intuition (i.e., if the other participant gives the rule-based answer) the extent to which people accept or reject such answer should more strongly depend on the perceived competence of the source. To evaluate such hypothesis, in the design of this second experiment, half of the other participant's answers were heuristic-based and incorrect answers, and the other half were rule-based and correct answers.

Furthermore, in order to strengthen the manipulation of other participant's competence, participants were asked to compare two profiles designed to be perceived as high (vs. low) in logical reasoning competence. These profiles were presented as describing previous participants in the study, and one of them was afterwards presented as the profile of the other participant with whom the participants were going to be paired.

Finally, the set of problems used was restricted to logical reasoning problems - syllogisms and transitive reasoning. By reducing the heterogeneity of the problems used we aimed to facilitate participants' learning processes.

Method

Participants. One-hundred sixty-one (140 females, M_{age} = 21.57, *SD* = 5.85) volunteers participated in the study. Part of the sample was composed of first-year psychology students (*N* = 99, 92 females, M_{age} = 19.10, *SD* = 4.01) who participated for a credit course. The remaining 62 participants (48 females, M_{age} = 25.50, *SD* = 6.19) were recruited from our lab subject pool using ORSEE (Greiner, 2015), and received a compensation of 10€ for their participation. Completing the study took about 25 minutes, and the study was part of a 1-hour experimental session that included further unrelated studies. Up to eight participants were recruited for each experimental session.

Materials.

Reasoning problems. Twenty-eight logical reasoning problems (14 syllogisms and 14 transitive reasoning problems) were distributed across the two phases of the experiment. In the first phase, the 24 problems were presented with an answer that was allegedly given by a previous participant in the study. Twenty of these twenty-four problems were presented in their conflict version (i.e., the believability of the conclusions opposes its logical validity); the remaining four problems (fillers) were presented in the no-conflict version (i.e., the believability of the conclusion is consistent with its logical validity). These fillers were included to have some trials in which the intuitive appealing answer is also the correct one, thus avoiding that participants adopted “the answer contrarily to intuition” as a response strategy. Half of the conflict version problems was followed by an heuristic-based answer while the other half was followed by a rule-based one. The four problems included in the second phase of the experiment were presented in their conflict version and without an answer from a previous participant. Two versions of material were created such that the problems presented with an heuristic-based answer in version A were presented with a rule-based answer version B and vice-versa (see Appendix C2.1. for the reasoning problems used in this experiment).

Syllogisms. Modus Tollens syllogisms were included together with the Modus Ponens syllogisms already used in Experiment 1. Hence, the syllogisms used were either logically valid - affirming the antecedent (Modus Ponens) or denying the consequent (Modus Tollens) or invalid - denying the antecedent or affirming the consequent. The syllogisms’ conclusions were either believable or unbelievable.

Transitive reasoning problems. Besides transitive reasoning problems involving one type of relation (as it happened in Experiment 1), problems involving two types of relations were also used. Thus, half of the problems presented premises with one type of relation (e.g., “Weeks are longer than minutes / Days are longer than minutes / (…)”) whereas the other half presented two types of relations (e.g., “Day is lighter than tibe / Night is darker than tibe / (…)”).

Cognitive Reflection Test. An adapted version of the extended CRT (S. Frederick, personal communication, February 2, 2016) with 6 items was used as an exploratory measure (see Appendix C2.1.). As in Experiment 1, different contents were used to control for prior exposure to the CRT and to avoid that participants could look up the answers on the internet. Participants recruited from our subject pool were presented with this test at the end of the experiment. The students completed the test in other experimental sessions, and a participant code was used to match data from the different sessions. Sixty-two participants were lost in the analyses that included this measure, due to difficulties in pairing our data with those from different experimental sessions.

Bias blind spot measure. Bias blind spot (BBS; Pronin et al., 2002) corresponds to the tendency to perceive others as more susceptible to biases than the self. In order to measure participants susceptibility to bias blind spot, participants were presented with the following paragraph (adapted from Pronin et al., 2002): 'Psychologists have claimed that, instead of thinking carefully about the problems they are presented, people often respond with the first thing that comes to their minds. As a result, people sometimes make errors in the way they reason and make judgments', and were then asked to rate (from 1 - *not at all*; to 7 - *strongly*) to what extent they thought that a) Portuguese students, b) Arts and Design students, c) Applied Mathematics and Computer Science students, and d) they themselves, tend to answer reasoning problems with the first answer that comes to mind rather then carefully thinking about the solution.

Other participant's profiles. Small modifications were made in the two profiles used in Experiment 1. Generically, participants described in the two profiles were more similar to those found in our lab's subject pool in terms of age and occupation. In the high competence condition, the other participant was presented as a 25 year-old male, generically described as 'Participant A', who was a PhD student in applied mathematics and computer science. His hobbies were reading sci-fi books, playing logic games, programming and tutoring math. In contrast, in the low competence condition the participant was a 20 year-old male, generically described as ‘Participant B’, who was studying arts and design. His hobbies were going out and drinking with friends, playing electro-trance music as a DJ, drawing tattoos, watching tv series and travelling by hitch-hiking (see Appendix C2.3. for the screenshots of profiles). The two profiles were presented as randomly chosen among those of previous participants who had authorized the researchers to share their answers.

Design. The experiment followed a 2 (other participant’s competence: low vs. high) x 2 (other participant’s answer: heuristic vs. rule-based) mixed design, with the first factor manipulated between participants.

Procedure. The experiment was presented as a study about how people solve logical reasoning problems in social contexts. The alleged other participant’s answers were presented as coming from a person who had previously participated in the study and authorized the researchers to share his/her answers. To make these instructions credible, before starting the experiment participants were asked to indicate if they authorized the sharing of their answers and, in affirmative case, they were asked to fill in some demographic information to be presented as their profiles to future participants (along with their answers).

To strengthen the manipulation of the other participant’s competence used in Experiment 1, before starting the experiment participants were presented with the two profiles and were asked to estimate each of these alleged participant’s performance on the logical reasoning problems (using a 7-points rating scale from 1 - *very bad*; to 7 - *very good*).

The study consisted of two phases. In the first phase, participants were presented with the 24 logical reasoning problems along with the answers given by the other alleged participant (see Materials section above). Their task was to indicate whether the conclusions of the logical reasoning problems were valid or invalid, stating how confident they were about their own answers.

Task instructions clarified what was meant by a 'valid conclusion' (i.e., a conclusion that logically followed from the premises, even if it violates one's prior beliefs), and all the premises and conclusions were explicitly identified in each trial (see Appendix C2.2. for the detailed instructions and an illustration of the trials).

The second phase of the experiment consisted of a new set of four logical reasoning problems similar to those found in the first phase, but this time presented without an answer from the other participant. Participants' task was to answer each reasoning problem, stating how confident they were about their answers, using a scale ranging from 1 - *not at all confident*, to 9 - *totally confident*. This second phase was included to test to what extent following or rejecting other participant's answers in the first phase would impact one's own reasoning.

At the end of the experiment participants were asked to complete the two measures of cognitive reflection and susceptibility to bias blind spot.

Results

Manipulation check. A t-test for independent samples was used to test the effectiveness of our manipulation of source competence. Results from this analysis showed that participants' initial estimates of other participant's performance were higher in the high (M = 6.20, SD = 0.78) than in the low competence condition (M = 4.06, SD = 1.04), $t(159) = 14.78$, $p < .001$, $d = 2.33$.

First phase problems.

Accuracy. Participants' accuracy was calculated as the proportion of correct answers after excluding the four no-conflict trials. The 2 x 2 ANOVA with other participant's competence (high vs. low) as between factor, other participant's answer (heuristic vs. rule-based) as a repeated measure revealed a main effect of other participant's answer, $F(1, 159) = 14.98$, $p < .001$, $\eta_p^2 = .09$. Participants' accuracy was higher when answering after rule-based (M = .76, 95% CI = [.73, .79]) than heuristic (M = .68, 95% CI = [.64, .72]) answers provided by the other participant. This main effect was qualified by a marginally significant interaction with other participant's competence, $F(1, 159) = 3.15$, $p = .078$, $\eta_p^2 = .02$ (see Figure 3).

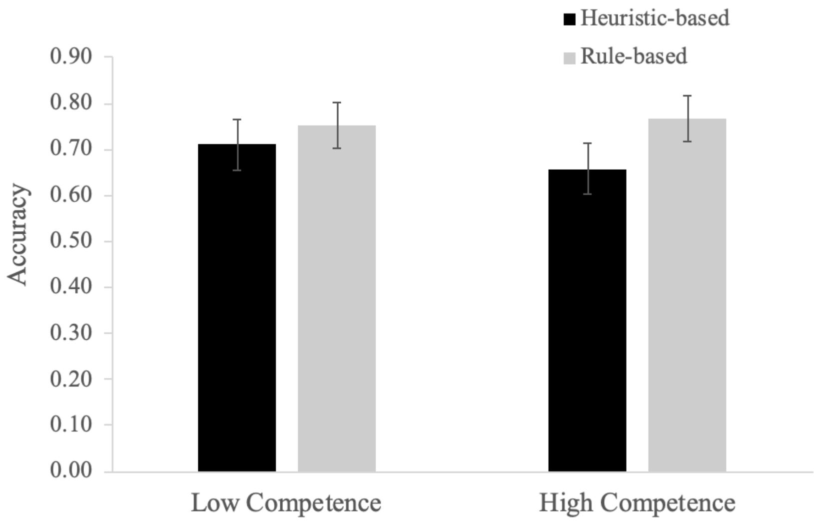


Figure 3. Proportion of correct answers by other participant's answer (heuristic vs rule-based) and other participant's competence (low vs high competence). Note: Error bars represent 95% confidence intervals of the mean.

While in the high competence condition participants' accuracy was higher after rule-based (M = .77, 95% CI = [.72, .82]) than heuristic answers (M = .66, 95% CI = [.60, .71]), $t(159) = 4.00$, $p < .001$, $d = 0.45$, in the low competence condition, no differences were found between participants' accuracy after rule-based (M = .75, 95% CI = [.70, .80]) and heuristic (M = .71, 95% CI = [.65, .77]) answers, $t(159) = 1.48$, $p = .142$, $d = 0.16$.

These results suggest that participants tended to follow other participant's answers to a greater extent when this other participant was presented as a high (vs. low) competent other. As in Experiment 1, participants' performance after other participant's heuristic answers did not differ across the experimental conditions (M_{HC} = .66, 95% CI = [.60, .71] vs. M_{LC} = .71, 95% CI = [.65, .77], $t(159) = 1.33$, $p = .186$, $d = 0.19$).

Confidence. Figure 4 shows participants' mean confidence on correct and incorrect answers provided after rule-based and heuristic answers from a high (vs. low) competent other.

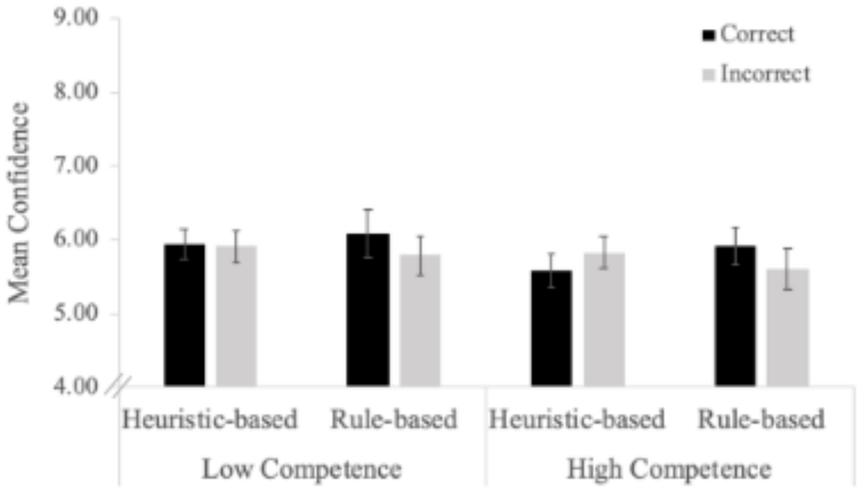


Figure 4. Mean confidence on correct and incorrect answers by other participant's answer (rule-based vs. heuristic) and other participant's competence (high vs. low competence). Note: Error bars represent 95% confidence intervals of the mean.

Results from a 2 x 2 x 2 ANOVA with other participant's competence (high vs. low) as a between factor, other participant's answer (rule-based vs. heuristic) and participants' own answer (rule-based and correct vs. heuristic and incorrect) as repeated measures revealed a significant Other participant's answer x Participants' own answer interaction, $F(1, 104) = 10.30$, $p = .002$, η_p^2=.090. When answering after other participant's rule-based answers, participants' confidence was higher for their rule-based and correct (M = 6.00, 95% CI = [5.85, 6.15]) than for their heuristic and incorrect (M = 5.69, 95% CI = [5.50, 5.89]) answers. On the contrary, when answering after other's heuristic answers, participants' confidence was higher for their heuristic and incorrect (*M* = 5.87, 95% CI = [5.66, 6.07]) than for rule-based and correct answers (*M* = 5.75, 95% CI = [5.60, 5.91]) answers.

Although the interaction Other participant's answer x Other participant's competence x Participants' own answer was not statistically significant, $F(1, 104) = 1.19$, $p = .278$, $\eta_p^2 = .01$, results for Other participant's heuristic answers replicated those found in Experiment 1. While in the HC condition participants were less confident on their rule-based and correct (*M* = 5.58, 95% CI = [5.37, 5.79]) than on their heuristic and incorrect (*M* = 5.83, 95% CI = [5.54, 6.11]) answers ($t(104) = 1.81$, $p = .037$ (unilateral), $d = 0.32$); in the LC condition, participants' confidence on their rule-based and correct (*M* = 5.93, 95% CI = [5.72, 6.15]) and heuristic and incorrect (*M* = 5.91, 95% CI = [5.62, 6.19]) answers did not differ ($t < 1$).

Second phase problems

Accuracy. Participants' accuracy on the second phase problems did not differ across the experimental conditions, $t(159) = 1.37$, $p = .172$, $d = 0.22$. However, participants in the LC condition improved their performance from the first (*M* = .73, *SD* = .20) to the second phase (*M* = .79, *SD* = .22, $t(79) = 2.73$, $p = .008$, $d = 0.30$). On the contrary, for those answering in the HC condition, no differences were found between performance in the first (*M* = .71, *SD* = .21) and second (*M* = .74, *SD* = .24) phase problems, $t(80) = 1.59$, $p = .116$, $d = 0.17$.

Participants paired with a low competent other were expected to more often second guess their answers and deliberate more. This could lead to enhanced accuracy which could transfer to the second phase problems. However, results do not support such possibility. As aforementioned, although participants paired with a low competent other showed an increase in performance from the first to the second phase there were no differences in accuracy across experimental conditions in either the first or second phase problems.

Additionally, we found a positive and significant correlation between participants' performance in the two phases of problems, $r(161) = .62$, $p < .001$.

Confidence. In contrast with the results from the first phase problems, no differences were found between participants' confidence on correct ($M = 5.91$, 95% CI = [5.73, 6.09]) and incorrect ($M = 5.97$, 95% CI = [5.73, 6.20]), $F < 1$.

Individual differences in following others' answers

An additional set of analyses was conducted to explore the impact of individual differences, such as participants' cognitive reflection (as measured by the CRT) or their susceptibility to bias blind spot, on the extent to which they follow (or reject) others' answers to the reasoning problems.

Influence measure. The degree of other participant's influence on participants' own answers was operationalized as the difference between their accuracy on problems presented with other's deliberate answers and the correspondent accuracy on problems presented with other's intuitive answers. This score could vary between -10 and 10, with negative values indicating a bias towards rejecting other's answers, and positive values indicating a bias towards accepting other's answers. The more participants discriminate between others' answers that should and should not be accepted the closer to zero the score in this influence measure should be.

Performance on CRT. Performance on CRT could vary between 0 and 6, with higher values revealing higher reflective thinking skills. Participants' mean performance on CRT did not differ across the experimental conditions, $t < 1$, with both conditions exhibiting relatively low levels of performance ($M_{LC} = 2.50$, $SD_{LC} = 1.79$ vs. $M_{HC} = 2.78$, $SD_{LC} = 1.80$). A median split of the CRT performance (Median = 2.00) was used to divide the sample in two groups: Low vs. high CRT.

Bias Blind Spot. The bias blind spot measure was calculated by subtracting participants' ratings of their own tendency to answer reasoning problems with the first answer that comes to mind instead of carefully thinking about the solution[12] from their ratings of the same tendency in other participant's group (i.e., Arts and Design and Applied Mathematics and Computer Science students). This measure could vary between -6 and 6, with negative values meaning that others were perceived as less prone than the self to answer reasoning problems intuitively and positive values meaning that others were perceived as more prone to answer intuitively than the self.

Results of the BBS measure showed that participants were evenly distributed across three groups: those who considered people from the other participant's group as less intuitive than themselves (BBS < 0; $N = 51$, $N_{HC} = 34$, $N_{LC} = 17$), those who considered themselves as intuitive as people from other participant's group (BBS = 0; $N = 59$, $N_{HC} = 33$, $N_{LC} = 26$), and those that considered themselves as less intuitive than people from other participant's group (BBS > 0; N = 51, $\mathrm{N_{HC}} = 14$, $\mathrm{N_{LC}} = 37$).

Following vs. rejecting other's answers. A 2 x 2 x 3 ANOVA with other participant's competence (low vs. high), cognitive reflection (low vs. high), and BBS (*more intuitive than others* vs. *as intuitive as others* vs. *less intuitive than others*) as between factors, and other's influence as the dependent measure revealed a main effect of cognitive reflection and a main effect of BBS. The main effect of cognitive reflection, $F(1, 87) = 3.94$, $p = .050$, $\eta_p^2 = .04$, showed that low CRT participants followed other's answers more (M = 1.23, 95% CI = [0.47, 1.99]) than high CRT participants ($M = 0.17$, 95% CI = [-0.57, 0.91]).

The main effect of BBS revealed significant differences across the three levels of BBS, $F(2, 87) = 4.54$, $p = .013$, $\eta_p^2 = .09$. Planned comparisons showed that the tendency to follow other's answers was higher for participants who consider themselves more intuitive than people from the other participant's group ($M = 1.84$, 95% CI = [0.92, 2.76]), than for the remaining

[12] Note that this tendency was presented as a source of many errors in people's reasoning processes.

two groups ($M = 0.13$, 95% CI =[-0.78, 1.04]), $t(87) = 3.01$, $p = .003$, $d = 0.65$). Participants in the as intuitive as others group ($M = 0.18$, 95% CI = [-0.68, 1.04]), followed other's answers to the same extent as those in the less intuitive than others group ($M = 0.08$, 95% CI = [-0.89, 1.04]), $t < 1$.

Although participants in the LC condition followed other's answers less ($M = 0.35$, 95% CI = [-0.39, 1.09]) than those in the HC condition ($M = 1.05$, 95% CI = [0.29, 1.81]), the main effect of other participant's competence did not reach statistical significance, $F(1, 87) = 1.72$, $p = .097$ (unilateral), $\eta_p^2 = .02$. Comparisons of the results from these two conditions with zero (the score that reflects an appropriate discrimination between answers that should be followed vs. rejected) revealed a significant bias towards following other's answer in the HC condition, $t(81) = 4.05$, $p < .001$, $d = 0.45$. Results from the LC condition did not differ from zero, $t(80) = 1.46$, $p = .148$, $d = 0.16$, suggesting that participants in this condition were more appropriately discriminating between answers to be followed and answers to be rejected. Additionally, the frequency with which participants were biased towards following (HC = 43 vs LC = 34) vs. rejecting (HC = 19 vs LC = 32) other's answers was significantly different across the experimental conditions, $X^2 (1, N = 128) = 4.24$, $p = .039$.

Figure 5 shows the mean influence of other participant's answers as a function of other participant's competence, performance on CRT (on the top) and BBS (on the bottom). Despite no interactions were found between the individual measures and other participant's competence (Fs = [0.10, 1.96], ps = [.165, .909]), some trends in the data are worth mentioning for future exploratory purposes.

First, participants' cognitive reflection influenced the extent to which they followed or rejected other participant's answers as a function of her competence. While low CRT participants tended to follow more the answers provided by a high ($M = 1.95$, 95% CI = [0.76, 3.14]) than a low ($M = 0.51$, 95% CI = [-0.44, 1.45]) competent other,

$t(87) = 1.89$, $p = .062$, $d = 0.54$, high CRT participants were not affected by Other's competence ($t < 1$).

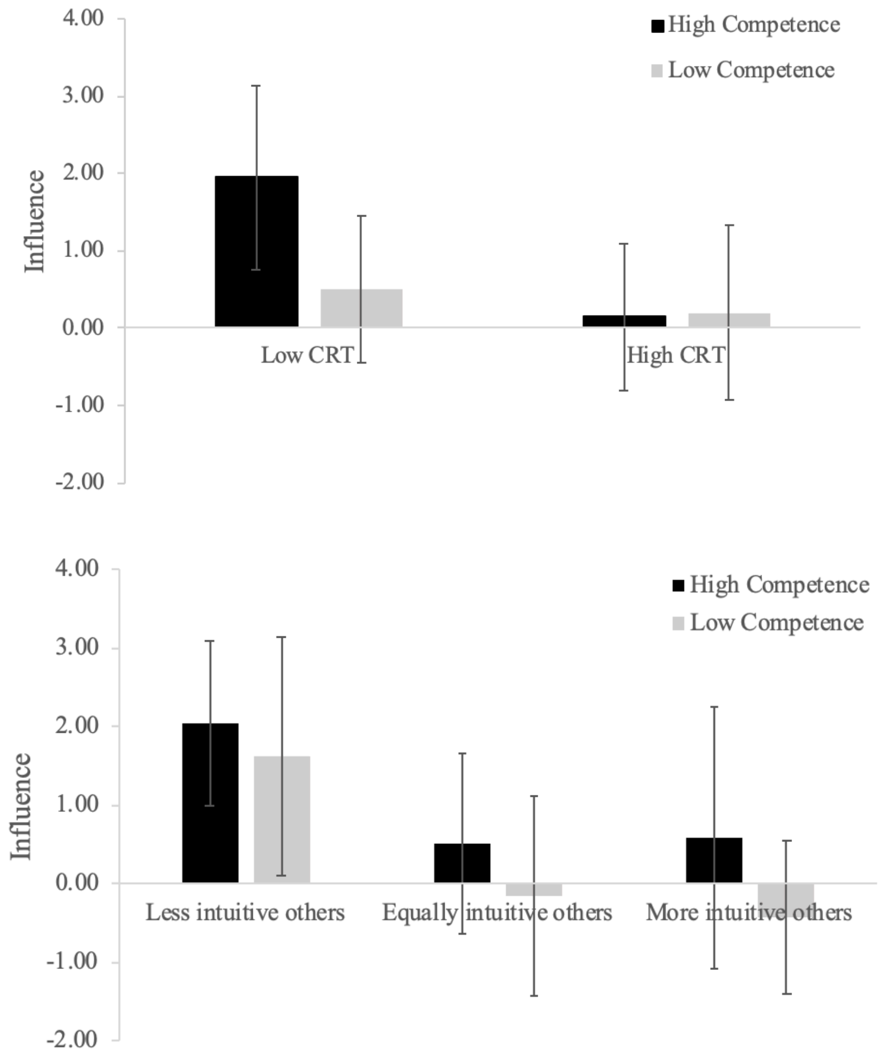


Figure 5. Mean influence of other participant's answers as a function of other participant's competence, performance on CRT (on the top) and BBS (on the bottom). Note: Error bars represent 95% confidence intervals of the mean.

The extent to which participants followed or rejected the answers provided by a high (vs. low) competence seems to have been differently affected by participants' level of BBS. In the LC condition, the tendency to follow other's answers was higher for participants in the 'more intuitive than others' group and linearly decreased across the 'as intuitive as others' and 'less intuitive than others' groups. On the contrary, participants' tendency to follow other's answers in the HC condition was less affected by their level of BBS.[13]

Correlation analyses revealed that the degree to which participants in the HC condition followed or rejected other's answers in the first phase problems, is negatively associated with their performance in the second phase problems, $r(81) = -.29$, $p = .008$. This same correlation was not significant in the LC condition, $r(80) = .05$, $p = .681$.

Confidence after following vs. rejecting other's answers. A 2 x 2 x 2 x 3 ANOVA was conducted to compare participants' confidence about their answers as a function of other participant's influence (following vs. rejecting other's answers), other participant's competence (low vs. high), participants' cognitive reflection (low vs. high), and participants' level of BBS (more intuitive than others vs. as intuitive as others vs. less intuitive than others), with all factors between participants except the first. Results from this analysis revealed no significant main effects nor two-way interactions, Fs = [0.01, 2.70], ps = [.104, .970].

Although the interaction Other participant's competence x Other participant's influence was not statistical significant, $F(1, 87) = 2.70$, $p = .104$, $\eta_p^2 = .03$, the pattern of results is consistent with the findings from Experiment 1. Participants in the HC condition tended to express lower levels of confidence when rejecting ($M = 5.92$, 95% CI = [5.65, 6.18]) than when

[13] We should note that these differences were statistically significant in the LC condition ($t(87) = 2.26$, $p = .027$, $d = 0.62$) but not in the HC condition ($t(87) = 1.48$, $p = .141$, $d = 0.76$). However, given that the groups involved in these comparisons were highly unbalanced, further research is needed to clarify these results.

following (M = 6.06, 95% CI = [5.80, 6.33]) other's answers, $t(87) = 1.46$, $p = .074$, $d = 0.20$. On the contrary, in the LC condition, participants' confidence about their answers did not differ as a function of following (M = 5.99, 95% CI = [5.73, 6.26]) or rejecting (M = 6.08, 95% CI = [5.82, 6.33]) other's answers, $t < 1$.

Discussion

As predicted, results from this experiment suggest that participants are more critical when evaluating answers coming from a low (vs. high) competent other. While in the HC condition, participants were more accurate when answering after other participant's rule-based (vs. heuristic) answers, this difference was not found for those answering in the LC condition. In the same vein, analysis using influence measure we created revealed a significant bias towards following other's answers in the HC but not in LC conditions.

Additionally, individual measures such as participants' cognitive reflection and susceptibility to bias blind spot were found to influence their tendency to follow other participant's answers, such that it is higher among participants with lower tendency to reflect (i.e. low CRT participants) and among those who consider themselves more intuitive than people from the other participant's group (i.e., participants less prone to show a BBS).

Replicating results from the previous experiment, the manipulation of other participant's competence had an impact on participants' confidence levels. When presented with heuristic answers from a high competent other, participants were less confident in their rule-based and correct than in their heuristic and incorrect answers. No differences were found when the same heuristic answers were provided by a low competence other. When other participant's answers were the rule-based ones, participants were more confident in their rule-based and correct than in their heuristic and incorrect answers, both in the LC and in the HC conditions. In sum, it seems that dissenting from other participant's answers had a metacognitive cost for participants when this other participant was described as high in competence and the answer provided was heuristic an intuitive appealing heuristic answer. The extent to which this "cost" might prevent participants from actually rejecting the other's answers will be further discussed in the General Discussion.

General Discussion

Most previous research on human judgment has focused on people's reasoning on decontextualized tasks or abstract problems, neglecting the fact most of our decisions take place in social settings, where people have access to other people's answers that might influence one's reasoning and decisions. The experiments included in this paper aimed to test whether people's performance in classic reasoning and decision-making tasks is (positively or negatively) affected when people can use others' answers to guide their own reasoning.

For that purpose, participants were asked to solve several classical reasoning tasks that present a conflict between heuristic- and rule-base answers. Before responding to each task participants were presented with an answer allegedly provided by other participants that were either working online with our participants (in Experiment 1) or that had previously participated in the study (Experiment 2). In both experiments, the competence of other participants was manipulated by using experimentally manipulated profiles of people stereotypically high (vs. low) on analytical competence. The answers allegedly given by the other participant were always the heuristic ones (Experiment 1) or a mixed set of intuitive and rule-based answers (Experiment 2). To explore carry over effects of the influence of others' answers in participants' own reasoning, both experiments ended with a new block of similar reasoning tasks that participants had to respond without having access to other's answers.

Our main prediction was that people would be more critical when analyzing the answers provided by others described as low (vs. high) in competence. In other words, profiles suggesting lower intellectual competence (compared to profiles high in intellectual competence) should make participants second guess the other's heuristic answers and engage in deliberate processing to confirm/overcome them. On the contrary, the same answers provided by a high competent other were expected to be followed without further analysis. Therefore, being exposed to heuristic answers from a low competent other should improve participants' accuracy on their own answers. Since this improvement is the result of additional deliberation, it was further expected this effect to carry over to a final set of problems where no answers from other participants were provided. That is, participants in the low competence condition should also reveal better performance in the second phase problems.

Contrary to our predictions, when heuristic answers were provided by others in Experiment 1, participants' accuracy was low and did not differ across the two other participant's competence conditions. This result suggests that once others' responses are made available participants behave as cognitive misers and uncritically adopt these responses. Indeed, when, in the second phase, we removed others' responses obliging participants to respond on their own, performance increased. The same pattern of results was found regardless of participants' level of cognitive reflection.

Although the manipulation of other participant's competence did not affect participants' accuracy in responding to the reasoning problems, it had an impact on their confidence levels. To give an accurate answer in this first experiment, participants had to reject the answer provided by the other participant. Remarkably, low CRT participants were less confident about their correct than incorrect answers when dissenting from the heuristic answers given by a highly competent other. In contrast, no differences were found between participants' confidence on correct and incorrect answers, suggesting that dissenting from a low competent other did not affect participants' confidence levels. The confidence levels of high CRT participants were not affected neither by the accuracy of their own answers not by the competence of the participant who had allegedly gave the answers provided with the reasoning problems.

Moreover, being exposed to heuristic answers that come from a highly competent source (compared to a source perceived to be low in competence) has negative lingering effects in terms of participants' metacognition. Indeed, when the second phase problems were presented, low CRT participants in the high competence condition remained less confident on their correct than incorrect answers although these problems were presented without answers from others. Again, no differences were found between confidence levels expressed by high CRT participants on their correct and incorrect answers.

As aforementioned, the fact that the other participant's answers were always the heuristic and highly intuitive appealing answers that most people give to this type of reasoning problems may have lowered participants' epistemic vigilance. In other words, given the absence of conflict between the answer provided and most participants' answers, participants had no strong reasons to further consider the profile information (i.e., the source of the answer) before responding. As a result, participants' accuracy was not affected by the stereotypically induced level of the other participant's competence.

To address this limitation, in Experiment 2 we presented participants with both heuristic and rule-based (and correct) answers provided by the other participant. In addition, Experiment 2 used a more clear-cut manipulation of other participant's competence, and only used logical reasoning problems.

As predicted, results from Experiment 2 showed that participants in the high competence condition followed other's answers more often than those in the low competence condition. As a consequence, participants in the high competence condition were more accurate when answering after rule-based than heuristic answers. This same difference was no found when participants were answering in the low competence condition, suggesting that the latter participants engaged more often in deliberate reasoning to correct the heuristic answers (and to confirm the rule-based answers) provided by the other participant.

Replicating results from Experiment 1, participants' confidence levels were affected by the manipulation of other participant's competence, such that participants became less confident when dissenting from a high competence (but not from a low competence) other - even when they were giving the correct answers (i.e., dissenting from other's intuitive answers). Remarkably, these results contrast with most previous research showing that correct answers (compared to the incorrect ones) are often accompanied by higher levels of confidence (e.g., De Neys, Cromheeke, & Osman, 2011; Mata & Almeida, 2014; Mata, Ferreira, & Sherman, 2013; Thompson, 2009; Thompson et al., 2011; but see Johnson-Laird & Savary, 1999), and pose some questions to the interpretations that have been advanced for those classical findings.

It has been suggested that higher confidence on correct answers reflect a metacognitive advantage that correct responders have, resultant from knowing that their answer is the right one, despite an alternative and highly intuitive answer exists (Mata, Ferreira & Sherman, 2013; see also Mata, 2019). The present results suggest that those feelings of confidence stemming from giving the correct deliberate answer may be more fragile than expected. Indeed, the metacognitive advantage that correct responders have when answering alone is weakened when, in a previous social context, they are presented with others' answers that diverge from their own.

Participants' confidence on correct and incorrect answers has also been used to test whether people detect the conflict between heuristic answers and the logical or statistical principles evoked in the problems typically used in this field (e.g., De Neys, 2014; De Neys et al., 2011). Specifically, the fact that people are less confident when giving incorrect answers to conflict problems than when giving the correct ones to no-conflict problems (in which heuristics and logical or statistical rules converge to the same answer) has been interpreted as a sign that people implicitly detect the abovementioned conflict, even if they cannot provide the correct answer.

The present experiments provide new evidence suggesting that confidence measures might not be a reliable measure of conflict detection when other sources of potential conflict are present in the decision context. In fact, when answering in social contexts, there is not only a potential intrapsychic conflict between the heuristic and rule-based answers but also an inter-psychic conflict between the answers provided by others and one's own answers. Interestingly, such inter-psychic conflict when it occurs seems to dominate the intra-psychic conflict[14]. More experiments should be conducted to better understand the interplay between these two types of conflict in people's reasoning and metacognition (e.g., whether people modify their answers as a consequence of experience the aforementioned inter-psychic conflict).

The present experiments may be seen as an extension of Mata, Fiedler et al.'s (2013) findings on the effects of being vigilant towards others' reasoning. While this previous research tested the effects of people's disposition to be vigilant, in the present experiments the vigilance towards others' reasoning was situationally induced by presenting participants with answers allegedly provided by high vs. low competence others.

Previous research aimed at studying reasoning within social-influence settings had already shown that exposing participants to the reasoning of others presented as less (but not

---

[14] We should note that we compared participants' confidence on correct and incorrect answers to conflict problems, instead of comparing their confidence on incorrect answers to conflict problems and correct answers to no-conflict problems. However, we believe that the same pattern of results should be found when using participants' confidence on correct answers to no-conflict problems as the baseline for the comparison. If anything, confidence levels for correct responses to conflict problems might be be lower than for correct responses to no-conflict problems (given the presence of a conflicting intuitive response in the first case but not the second).

as more) competent improves individuals' reasoning in hypothesis testing tasks by increasing the use of disconfirmatory strategies (Butera et al., 2005; Butera & Mugny, 1995). These results were framed within the conflict elaboration theory of social influence (Mugny, Butera, Sanchez-Mazas, & Perez, 1995; Pérez & Mugny, 1993, 1996). According to this theory, being exposed to an answer that is different from one's own creates a conflict which resolution depends on the source of the answer and the type of the tasks. In problem-solving tasks, people are motivated to seek the correct answer as their aptitude is at stake. For this reason, divergent answers provided by a competent source are interpreted as informational support, and are therefore imitated with little processing of the task. On the contrary, divergent answers provided by a less competent source lead to a conflict: on one hand, people are not certain about the correctness of their own answer, on the other hand, they do not want to accept the source's answer as the likelihood of being correct is perceived as low. This so-called conflict of incompetences leads to a deeper processing of the task, which eventually results in better performance.

Divergence from others thus seems to be crucial to make improvements in one's reasoning. This idea is also present on research on epistemic vigilance, according to which people's vigilance towards others depends on the coherence/similarity between the information they provide and people's background information/knowledge (coherence checking principle) (Sperber et al., 2010). Some empirical evidence can also be found in advice-taking research that shows that people's use of others' advice is influenced by the similarity between the advice and people's own judgments (Wanzel, Schultze, & Schulz-Hardt, 2017; Yaniv, 2004; Yaniv & Milyavsky, 2007).

It may be insightful to look at the present experiments results from this perspective. According to these theories and conceptual approaches, when dealing with the type of reasoning problems used here, people are vigilant towards the information contents (i.e., if the answers provided by others agreed with their own beliefs/answers) and towards who provides the information (i.e., whether the others are competent to solve the tasks or not). It follows that when other's answers are coherent with previous beliefs/their own, people are more likely to accept them at face value, lowering their vigilance towards who provided the answers (i.e., less likely to take into further consideration the competence of the source). Given that, in Experiment 1, the other participant's answers were always heuristic-based and since such answers correspond to the intuitively appealing answers that come to mind to most people, it is likely that participants have, for the most part, accepted these other's answers without taking into further consideration the other's competence level. In contrast, when the other participant provides rule-based and hence counterintuitive answers that conflict with most participants' answers, then in order to decide if they should reject these answers (and stick to their intuitive heuristic ones) or accept them, they are more likely to be vigilant towards the source of the (rule-based) answers, taking into consideration the profile of the other respondent. Furthermore, such cognitive processes of epistemic vigilance are likely to engage participants in effortful, deliberative reasoning. As such, when accepting rule-based answers, participants were not merely following the other in an uncritically way, they are revising and adjusting their previous (heuristic-based) beliefs.

Limitations and future studies. Our results are just a starting point to test these hypotheses, and should be interpreted with caution due to their low effect sizes. In fact, when all the answers presented to our participants were the heuristic and intuitively appealing ones (Experiment 1), the competence attributed to source of the answers had no impact on the likelihood of accepting or rejecting them. As we had previously stated, this result can be partially attributed to some aspects of the experimental procedure. However, the fact that these answers are similar to those that people usually give should also have contributed to this result.

When heuristic answers were mixed with the rule-based ones (Experiment 2), participants engaged in deliberate reasoning more often when answering after a low competence source, while uncritically following the same answers when provided by a high competence source. This result seems to suggest that presenting people with answers that sometimes are similar but other times different to their own contributes to make them vigilant towards others' answers, and to trigger the processes associated to the exposure to different social influence sources.

To further explore the underlying processes of our results, future studies should measure participants' response times when answering after heuristic and rule-based answers from others, particularly when the type of answer provided is manipulated in between-subjects designs. Along the lines previously stated, heuristic answers from others should be accepted faster than the rule-based ones, and the competence of the source should not impact participants' response times. On the contrary, it should take longer to accept rule-based answers provided by a low competence source than the same answers provided by a high competence source.

Additionally, to disentangle whether the better performance found in the second block of problems is due to not having access to other participant's incorrect answers (supposedly accepted without scrutiny in the first block) or a mere practice effect, future studies should include a control condition, in which participants answer to both blocks of problems without having access to other participant's answers.

Future studies should also manipulate the type of answer provided by others in a between-subjects design in order to better understand the interaction effects of the answers' (dis)similarity and the competence of the source. To better explore the consequences that uncritically following vs. accepting others' answers might have for one's own reasoning, these future studies should include a final block of problems presented without the answers provided by others (similarly to what was done in the present experiments).

Moreover, to disentangle the impact of answers' (dis)similarity and its (in)correctness, future studies could also use problems that allow for the use of nonintuitive incorrect answers (i.e., answers that are incorrect but not systematically given by most people). Although the problems used in the current experiments do not allow for this kind of answers (except for disjunctive reasoning problems used in Experiment 1), there are some reasoning tasks typically used in the heuristics and biases research that allow for this manipulation (e.g., conjunction fallacy, base-rates neglect).

The impact of answers' (dis)similarity should also be explored with the manipulation of other characteristics of the source besides competence. Sperber et al. (2010) argue that people calibrate their trust on others according to their perceived competence and benevolence. It could be the case that people further scrutinize other people's answers when they have reasons to suspect of their motivations, even when these answers are similar to their own.

Future studies should also explore the consequences that being exposed to others' answers have for future interactions with the same source. Our results showed that dissenting from a high competence source lowered participants' confidence in their answers, even when they are correct. This same pattern spillover to a second block of trials in which participants no longer had the answers from others. It would be relevant to test the consequences of this lowered confidence in a new block of trials, in which participants would be answering to the participant that had given the answers provided in the previous block of trials. It could be the case that the lowered confidence found after correctly rejecting heuristic answers from a high competent other lead people to regress to heuristic answers when answering to this high competent other in a new block of trials.

Previous research aimed at better understanding what conditions make people recognize the output of T1 as potentially faulty and trigger the engagement of T2 has focused on the characteristics of the decision-maker (e.g.,; De Neys, 2006; Frederick, 2005; Nisbett et al., 1983; Stanovich, 1999, 2011; Thompson et al., 2011), and 2) characteristics of the task (e.g., Daniel & Klaczynski, 2006; Evans & Curtis-Holmes, 2005; Finucane et al., 2000; Vadenoncoeur & Markovits, 1999). The present experiments extended this research by testing the impact that decision environments might play in people's reliance on T1 or T2 processing. It was suggested and found that the exposure to other people's reasoning could affect the extent to which people correct (or on the contrary reinforce) their judgmental biases through the engagement of T2 processes. We should note, however, that the engagement of T2 processes does not warrant, per se, the correction of judgmental biases (e.g., De Neys, 2012; Evans, 2019; Evans & Stanovich, 2013; Pennycook, Fugelsang, & Koehler, 2015; Thompson, 2009). The engagement of T2 processes will only lead to correct answers when people have knowledge about the relevant rules and cognitive capacity to apply them (Evans & Stanovich, 2013; Stanovich, West, & Toplak, 2016). It follows that a more complete understanding of people's judgment and decision-making processes requires to further explore the combined effect that the decision environments, the individual characteristics of the decision-maker and the characteristics of the decision tasks (notwithstanding the importance of studying these aspects in isolation). The exploratory measures included in this paper might be viewed as a starting point for this research. Although the manipulations of other participant's competence were effective in cueing people to second guess their answers (particularly when the low competent other provided the rule-based answers for half of the trials), the impact of such manipulations was qualified by participants' individual characteristics such as their cognitive reflection.

Chapter 5 - General Discussion

The aim of this thesis was to empirically test the role that decision environments might play in one's reasoning processes. More specifically, I was interested in testing whether the errors and biases typically found in the HB literature would be corrected, attenuated or even reinforced depending on the extent to which decision environments cued people to second guess their intuitive answers or strategies.

The starting point of this project was the idea that errors and biases found in one-shot tasks may result from processes that are adaptive and functional in continuous and socially enriched contexts (Hogarth, 1981; see also Einhorn & Hogarth, 1978; 2001). In fact, by focusing on people's biased answers to discrete inferential tasks, most previous research in the HB tradition left unexplored how these answers unfold when people have the opportunity to progressively adjust their judgmental strategies in accordance to the feedback (either social or not) received from the decision environment (Brehmer, 1996; Hogarth, 1981; Kleinmuntz, 1985).

In contrast with previous research, the empirical work presented in this thesis built on the assumption that the flexibility of people's cognitive system should allow them to adjust their reliance on largely autonomous (T1) processes and more deliberate, rule-governed (T2) processing, according to the feedback received from the environment (Ferreira et al., 2006). In other words, the faulty answers commonly found in the HB one-shot tasks (mostly attributed to T1 processing) can be corrected or, on the contrary, reinforced depending on the extent to which decision environments signal the need to engage in more effortful T2 processing.

Two specific aspects of the decision environments were explored for their potential to correct or reinforce erroneous judgments (through a greater or lesser engagement of T2 processing). On one hand, using a classical pseudodiagnosticity task in a repeat-play experimental setting, it was explored to what extent the feedback provided in continuous learning contexts helps people to become aware of their errors or the inadequacy of their strategies (i.e., when accurate and timely feedback is provided in kind environments), thus triggering the engagement of more deliberated (T2) processing; or, on the contrary, reinforces people's reliance on T1 processing by keeping them inappropriately convinced about their adequacy (i.e., when misleading or incomplete feedback is provided in wicked environments) (Chapter 2).

On the other hand, using a modified version of the standard anchoring paradigm (Chapter 3) and classical reasoning problems presenting a conflict between heuristic-based and rule-based (logic or probabilistic) responses (Chapter 4) it was tested the role that other people might play in triggering T2 processing when people answer in the socially enriched contexts. It was hypothesized that other people's answers would trigger the engagement of T2 processing when people have reasons to be suspicious about these answers. That is, when these answers were provided by a low in competence or biased other and diverge from the answers that people would give on their own. On the contrary, when others' answers were provided by a highly competent other or when they are similar to people's own answers (regardless of the competence of the source), people would follow them uncritically (relying mostly on T1 processing) (Chapters 3 and 4).

Next, I summarize the main results of these sets of experiments along with a discussion of their implications. I firstly present and discuss the findings from experiments focused on the impact of feedback provided in continuous environments (Chapter 2), and then I turn to the experiments that explored the social dimension of decision environments (Chapters 3 and 4). The limitations of the present work and some ideas to overcome them in future research are presented throughout the discussion of the main findings of each set of experiments. Finally, I discuss contributions of the current findings to the theories of JDM research.

Human Judgment in Continuous Learning Contexts

Pseudodiagnosticity in continuous learning contexts

The first set of experiments (Chapter 2) tested whether and how people progressively adjust their judgmental strategies according to the feedback received when answering in continuous learning contexts. Using a multiple trial version of the pseudodiagnosticity paradigm (Doherty et al., 1979; 1981), these experiments examined how people's tendency to make pseudodiagnostic choices would evolve in decision environments designed to clearly reveal (kind environment) or mask (wicked environment) the inadequacy of participants' strategies. The prediction was that the use of pseudodiagnostic strategies decrease as people become aware of the inadequacy of these strategies - something that kind environments were expected to promote and wicked environments to hinder.

Results from two experiments showed that although participants progressively made more diagnostic choices in both environments, improvement in performance was further promoted by kind environments. In contrast to Experiment 1, a kind decision environment used in Experiment 2 did not significantly improved performance (when compared with the wicked environment). However, participants' performance on a new and more complex decision scenario included at the end of the experiment indicates that answering in kind environments actually improved participants' deep understanding of the task, whereas the performance improvement acquired in wicked environment was more superficial and did not transfer to the new decision scenario.

Taken together, these results provide some support to the idea that judgmental errors and biases found in one-shot experiments could disappear (or at least be attenuated) in continuous learning contexts, where people might become aware of their errors and have opportunities to adjust their judgmental strategies accordingly.

As aforementioned, this research was inspired by Hogarth's (2001; Hogarth et al., 2015; Hogarth & Soyer, 2011) seminal work on how the structure of decision environments conditions the quality of what people learn in the course of their lives. From this perspective, many potentially dysfunctional strategies reported in the JDM literature reflect habits that people have gained in their natural decision environments, where the available information is limited and feedback is often missing of misleading (see also Einhorn & Hogarth, 1978). To counteract the effects of these rooted habits, Hogarth (2001; see also Hogarth et al., 2015; Hogarth & Soyer, 2011) argued that people need to deliberately seek or create kind decision environments, where feedback is accurate and helpful to correct their erroneous judgments.

The present proposal complements this previous research by testing whether decision environments could facilitate (or difficult) the detection and override of the highly appealing pseudodiagnostic choice (T1) by signaling the need to second guess such initial hypothesis-testing intuition and engage in more elaborate (T2) testing strategies.

In a sense, this research may also be viewed as a tentative answer to the often made criticism that HB research program used experimental paradigms and materials purposefully created to make people err (e.g., Gigerenzer et al., 1999; Goldstein & Gigerenzer, 2008; Hertwig & Ortmann, 2007; Todd & Gigerenzer, 2003). From Gigerenzer and collaborators' perspective, people's intuitive answers were developed through human evolution in their natural environments and, for this reason, they are right most of the time. The errors found in HB tradition, they claim, result from choosing specific examples in which people's intuitions do not work. Notwithstanding the merits of this criticism (in fact, in many occasions people's intuitions are highly efficient and lead to the correct answers), these authors may have neglected the fact that this same distortion in the ecological validity of the decision environments could be made by individuals interested in exploiting others' intuitive answers for their own advantage (Evans & Stanovich, 2013; Stanovich, 2018). Therefore, when people are answering in ecologically invalid environments, it is crucial to better understand what makes them reject their heuristic-based (T1) answers and engage in more elaborate T2 reasoning processes.

The optimistic message to be taken from the findings here reported is that judgmental errors and biases found in discrete tasks can at least be attenuated in decision environments that give people the opportunity to become aware of their errors or the inadequacy of their judgmental strategies. Not so optimistic, however, is the fact that many of our real-world decision environments fail to provide people with these opportunities, thus hindering the correction of judgmental errors and biases.

To avoid the detrimental effects of answering in wicked environments, Hogarth (2001; Hogarth et al., 2015) suggested that people (and eventually policy makers) should learn to recognize the limits of the experience provided by the (social) environments in which decisions are usually made and deliberately seek or create more kind environments. The impact of such kind environments on people's reasoning processes has been empirically tested in a set of experiments using behavioral simulations (Hogarth & Soyer, 2011; Soyer & Hogarth, 2015). Instead of presenting participants with nontransparent descriptions of problems (as it is usually the case in experiments in the HB field), these behavioral simulations allow participants to directly experience the outcomes of the processes about which they are asked to make inferences, thus facilitating the correction of judgmental biases (see also Hertwig & Erev, 2009; Hertwig et al., 2018; Wulff, Mergenthaler-Canseco, & Hertwig, 2018 on the description-experience gap).

From my perspective, such behavioral simulations are conceptually similar to the experimental procedure used in the present experiments in that both let people experience the consequences of their actions or decisions. However, whereas this previous research focused on people's experience in kind environments, where people receive accurate and timely feedback that helps them to recognize the need to make eventual adjustments in the judgmental strategies used, the present experiments extended their focus to also test the impact of making judgments in wicked environments, where missing or misleading feedback might keep people unaware of the need to correct their erroneous judgments. Rather than trigger the correction of judgmental biases, these so-called wicked environments might even reinforce the use of inadequate judgmental strategies. Results from the present experiments do not entirely support this prediction in that both the wicked and the kind environments led people to reduce their pseudodiagnostic choices throughout the trials, though to a lesser extent in the former ones.

Two aspects of these results are worth discussing here. On one hand, the fact that people improve their performance when answering in continuous learning contexts, suggests that one-shot experiments that are typically used in the HB research tradition might be insufficient to derive definite conclusions about the functioning of human judgment processes. In fact, judgmental biases found with one-shot experiments seem to be attenuated when people are given more opportunities to answer similar problems. On the other hand, it might be the case that the wicked environments used were not as wicked and it was intended. In fact, the pseudodiagnosticity paradigm used in these experiments is quite transparent regarding the structure of task, which might have facilitated participants' learning and reduced the strength of the manipulation used. Therefore, the impact that wicked environments might have in the extent to which people modify or maintain their judgmental strategies should continue to be tested in the future using less transparent paradigms or, in Hogarth and colleagues' (2015) terms, using decision environments that vary in their position in the continuum between a totally kind and a totally wicked environment. In this regard, further experiments could test, for instance, whether kind (vs. wicked) decision environments facilitate (or hinder) people's awareness of the need to adjust their estimates of uncertain quantities (i.e., anchoring task) or the weight they give to base rate and case-specific information in their probabilistic predictions (i.e., base rate task).

In the present experiments, the wickedness (vs. kindness) of decision environments was manipulated by presenting participants with information consistent (vs. inconsistent) with the one they already expected; and by making the normatively correct and incorrect answers to converge (vs. or not converge) in the same answer. However, other features of the decision environments might be used to create kind and wicked environments. Given that what makes an environment kind (vs. wicked) is the extent to which it allows people to be aware of their errors or the inadequacy of their judgmental strategies, these two types of decision environments might also be defined, among other ways, by their tolerance to error (i.e., how strict (kind) vs. lenient (wicked) is the range of answers they accept as correct).

To illustrate, consider the case of the classic two-steps anchoring paradigm in which participants are asked to make an absolute numeric estimate about some property of a given entity after considering whether it is lower or higher than a value provided as an anchor (e.g., estimating the distance between Baltimore and Chicago - i.e., 974km - after considering if it is longer or shorter than 2000km). Using the anchor value (i.e., 2000 km) as a starting point for adjustment, imagine that your answer was 1400km. Your answer would be considered correct in a lenient (and wicked) environment that accepted answers in the interval 474-1474km (i.e., answers that depart from the correct answer less than 500km), but it would be considered incorrect in a stricter environment that restricted the correct answers to the interval 874-1074km (i.e., answers that depart from the correct answer less than 100km). Although your answer was the same in both decision environments, when answering in the lenient-wicked environment it would be accepted as correct, therefore stopping/reducing your need to engage in more effortful processing. On the contrary, when answering in the strict-kind environment the same answer would be considered incorrect, making you aware of the need of further adjustments.

Along these lines, it is reasonable to assume that the stricter an environment is in the range of answers it accepts as correct, the more it provides people with opportunities for learning about their errors and, as a consequence, the more it triggers the engagement of correction procedures. However, it could also be the case that environments that are too strict or too lenient in the range of acceptable answers are both inappropriate for learning, though for different reasons. Specifically, always receiving positive feedback in too lenient environments might keep people unwarrantedly relying on the first answers that come their mind, not adjusting away from the values presented as anchors. On the other hand, systematically receiving negative feedback in too strict environments could reduce people's motivation to appropriately answer the task, making them give up on their debiasing efforts. In this regard, participants' performance should be better when answering in environments that are intermediate in their levels of tolerance to error, and equally poor/lower when answering in too strict or too lenient environments.

In fact, this latter prediction is consistent with Hogarth, Gibbs, McKenzie, and Marquis's (1991) results showing that participants' performance follows an inverted-U-shaped function of environment's tolerance to error, with higher performance being associated with intermediate levels of tolerance to error, and similar lower performance associated with both strict and lenient environments. Remarkably, the authors explained this pattern of results by the way people react to feedback: whereas positive feedback (more often received in lenient environments) lead people to maintain using the same strategies, negative feedback (more often associated with strict environments) encourages the search for alternative and potentially more appropriate strategies. Although the search for alternative strategies is crucial for the correction of judgmental biases, it seems that there is a limit upon which the inconsistency in responses induced by these alternative strategies becomes more detrimental than beneficial. Future experiments should further explore this issue using the classical HB tasks, in order to better understand what is the most appropriate level of tolerance to error to improve people's judgments in these tasks.

Future experiments should also test how human judgment adapts to changes in more dynamic decision environments than the ones studied here. Specifically, to what extent decision strategies will be updated when the environment change in such a way that makes these (previously adaptive) strategies become ecologically invalid. Previous research suggests that in the face of sudden changes people tend to persist in the old (no longer adaptive) strategies (e.g., Bröder, Glöckner, Betsch, Link, & Ettlin, 2013; Hogarth, 2001; Rieskamp & Otto, 2006; Schwartz, 1982) . However, environments that make more transparent the relevant changes in their ecology (and the consequences of such changes for judgment calibration) might be able to signal the metacognitive need for replacement of old strategies. The specific features of such dynamic (and yet) kind environments and how they interact with people's individual differences (in cultural knowledge, reasoning ability, motivation) are far from being fully understood and thus await further research.

Human Judgment in Social Contexts

Although people often rely on the information provided by others to make their decisions, this social dimension of decision environments has been largely unexplored in previous research in the HB tradition (Larrick, 2016). In order to fulfill such research gap, the two remaining sets of experiments tested the role that other people might play in the extent to which people correct (or on the contrary reinforce) their judgmental biases, through a greater (or lesser) engagement of T2 processes.

Inspired by previous research on the conflict elaboration theory of social influence (Mugny et al., 1995; Pérez & Mugny, 1993, 1996) and on Sperber et al.'s (2010; Mercier & Landemore, 2012; Mercier & Sperber, 2011, 2017) notion of epistemic vigilance, these experiments tested the idea that the information provided in social contexts might be further scrutinized (or, on the contrary, uncritically accepted) depending on the trustworthiness of the source (i.e., its competence and/or benevolence), and on how divergent it is from people's own knowledge and/or beliefs. Along these lines, it was proposed that being exposed to other people's answers would trigger the engagement of effortful T2 processes whenever participants had reasons to be suspicious about these others' answers (either because they diverge from one's own answers, or because they were provided by others perceived as biased or less competent).

To test this hypothesis, the experiments included in Chapter 3 explored whether people's tendency to perceive others as more biased than the self (Pronin, 2007; Pronin et al., 2002, 2004) make them more critical (and therefore more prone to engage in T2 processing) towards the information provided by others (who are potentially biased) (vs. the same information presented without a specified source). The experiments included in Chapter 4, specifically tested whether the extent to which people engage in T2 processing when presented with others' answers is affected by the competence of these others and by the similarity between their answers and those that most people would give on their own.

Anchoring in a social context

The experiments included in Chapter 3 used a modified version of the anchoring paradigm to test how people's naïve theories of biases (Ehrlinger et al., 2005; Pronin, 2007; Pronin et al., 2002, 2004; Wilson & Brekke, 1994; Wilson et al., 1996) impact the extent to which they engage in more elaborate, T2 processing when presented with estimates (i.e., anchor values) provided by others compared to the same estimates presented without a specified source.

These experiments built on the idea that previous failed attempts to reduce the anchoring bias through forewarnings result from people's pervasive tendency to deny their own biases while imputing them to others (Pronin, 2007; Pronin et al, 2002, 2004). Given this so-called bias blind spot, it was suggested that forewarnings about the anchoring bias would convince participants of others' (but not their own) susceptibility to these biases. Therefore, forewarnings would make participants treat others' estimates as biased, triggering debiasing efforts. In other words, when presented with estimates provided by others (vs. the same estimates presented without a specified source), participants who were forewarned about the anchoring bias, were expected to adjust their own estimates and show a reduced anchoring effect. Given that adjusting away from others' answers is an effortful process that takes time, these participants were further expected to take longer to give their answers.

This hypothesis was tested in two experiments that combined forewarnings about the anchoring effect with anchor values presented as other participants' answers or without a specified source. As predicted, results showed that combining forewarnings with anchors provided by a social source effectively reduced the anchoring effect. Furthermore, the response-time analysis in Experiment 1 suggested that such attenuation was the result of deliberate adjustment from the provided answers.

Overall, these findings provide further evidence for recent alternative explanations of the anchoring effect according to which externally provided anchors trigger the same adjustment (elaborate T2 processes) that were previously associated with self-generated anchors (Dowd et al., 2014; Epley & Gilovich, 2001, 2005; Simmons et al., 2010). In the present experiments, it was suggested and found that people's tendency to perceive others as more prone to bias than themselves (i.e., the bias blind spot; Pronin et al., 2002) would trigger adjustment processes whenever the anchor values were suggested by other people, and therefore perceived as biased. Put it differently, it was people's bias blind spot (which is also a bias) that made them further elaborate on others' answers, thus leading to a correction of their own biases. In this regard, the present results give support to the idea that the errors produced by the use of some heuristics might be compensated by others. To illustrate this idea, Krueger (2012) cited an unpublished experiment by Gideon Goldin and Leonard Chen (Brown University) in which the anchoring bias was found to reduce participants' base rate neglect in the lawyer-engineer problem (Kahneman & Tversky, 1973). In this classic base-rate problem participants are asked to estimate the probability that a personality description belongs to an engineer, given that it was drawn from a sample with a higher (vs. lower) percentage of engineers (and a complementary percentage of lawyers). When the base rates were used as an anchor (i.e., when participants were asked to consider whether the probability that the personality description belongs to an engineer is higher or lower than the base rates), participants' answers became anchored to these values, which means that base rates were no longer neglected.

Therefore, it seems important that future experiments explore the combined effects of using different heuristics rather than exploring them in isolation. In fact, it is possible that many biases found in real world settings result from the interaction of several heuristics and not from their isolated operation but it is also likely that the interaction or cumulative effect of different biases that work in different directions may off set or reduce people response bias (see Krueger & Funder, 2004 for illustrative examples). The specific circumstances where biases are amplified or reduced by the composite work of different heuristics are yet poorly understood and thus an interesting avenue for future research (for similar exemplars on the combined effect of heuristics regarding confirmation biases see Klayman, 1995; Nickerson, 1998).

At a broader level, these results are aligned with Sperber et al.'s, (2010; Mercier & Sperber, 2011) idea that people are vigilant towards the information provided by others in order to avoid being incidentally or deliberately misinformed by them. This epistemic vigilance towards others lead people to more likely detect flaws in their answers and reasoning, and thus trigger the engagement of elaborate T2 processing. This should be particularly the case when these others are perceived as low in competence (or in benevolence), as might be the case when forewarnings about biases are provided.

At this point, it is worth mentioning one methodological limitation of the present experiments. It was proposed that participants' bias blind spot would make them treat others' answers as biased (when forewarned about the anchoring effect), thus leading to a reduction in the anchoring effect. Although previous research has shown that BBS is quite prevalent in the population (Ehrlinger et al., 2005; Pronin, 2007; Pronin et al., 2002; 2004; West, Meserve, & Stanovich, 2012), none of the present experiments directly measured participants' BBS. The lack of this measure might have created some noise in the results, as some participants may not exhibit BBS (which is quite uncommon in the USA but not so uncommon in the experiments conducted in Europe, e.g., Mata, Fiedler et al., 2013). Indeed, participants who thought they were more biased than others (reversed-BBS participants) may have followed others' answers more often, eventually attenuating the reduction in the anchoring bias achieved by BBS participants. To clarify the role of BBS as a potential moderator of the reduction of anchoring effects in social contexts, future studies should include an individual measure of BBS.

Lastly, the anchor values used in these experiments were always quite infrequent (correspondent to percentiles 15th an 85th of the calibration group), following the common practice in the literature of anchoring effects (Jacowitz & Kahneman, 1995). However, to better explore how people treat others' answers as a function of their perceived trustworthiness (either resultant from people's BBS or from experimental manipulations of others' competence or benevolence), future experiments should also include anchor values more similar to the median estimate of the calibration group. Such values would allow to test whether people adopt a new heuristic strategy to answer the task (e.g., always adjusting from others' answers) or whether they consider others' answers as potentially correct, only adjusting away from those that seem too extreme (i.e., 15^{th} and 85^{th} percentiles).

Reasoning about others' heuristic- and rule-based answers

The experiments included in Chapter 4 extend those in Chapter 3 by exploring other ways in which the social dimension of decision environments might prompt people to doubt their intuitive answers and trigger the engagement of more elaborate T2 processing. Specifically, the two experiments tested whether being exposed to the answers provided by other participants described as high vs. low in competence would lead participants to correct their own judgments. In both experiments, participants were asked to solve several logical reasoning and decision-making tasks after considering the answers allegedly provided by other participants in the study. In both experiments, these other participants were described as stereotypically high (vs. low) on analytical competence. The answers provided by others were either the heuristic but incorrect answers that most people give to these reasoning problems (Experiment 1) or a mixed set of heuristic and rule-based correct answers (Experiment 2). Both experiments included a final block of problems to test whether the cognitive effects of reasoning about others' reasoning would spill over to subsequent reasoning problems this time presented without answers from other participants. Individual measures of participants' cognitive reflection (Experiments 1 and 2) and bias blind spot (Experiment 2) were included to explore their potential impact in the extent to which participants follow or reject others' answers.

The results showed that the competence of other participants alone was not sufficient to make people more (vs. less) critical towards their answers. In fact, when participants were presented with the heuristic and highly intuitive answers allegedly provided by other participant in the study, no differences were found between those answering in the low and in the high competence conditions (Experiment 1). The low levels of accuracy found in both experimental conditions suggest that participants uncritically followed the answers provided by others. On the contrary, participants were more critical towards the answers provided by a low (vs. high) competent other when these answers were a mixed set of heuristic- and rule-based answers (Experiment 2). Specifically, participants in the high competence condition still followed other's responses and thus were more accurate when answering after other's valid answers and less accurate after other's heuristic answers. Participants in the low competence condition reflected more about other's answers, rejecting heuristic answers and accepting valid answers. In the same vein, results from the influence measure revealed a significant bias towards following other's answers in the first but not in latter condition. The impact of the abovementioned manipulations mainly affected participants with low cognitive reflection (Experiments 1 and 2), and participants who believed that others are less intuitive than themselves (Experiment 2). Specifically, participants low in cognitive reflection were more influenced by the competence of the source (i.e., they more often followed the answers provided by a high than a low competent other). In contrast, responses of participants high in cognitive reflection, were less affected by the competence of the source (i.e., their answers did not significantly differ across the two experimental conditions: high and low competence of the other). In a similar vein, participants who believed that others are less intuitive (and, as a consequence, less biased) than themselves, tended to follow other participant's answers more often than those who believed that others are as intuitive or even more intuitive than themselves.

Taken together, these results suggest that decisions to engage in deeper deliberate reasoning (T2) follows a three-step metacognitive process. As long as others' answers converge with most participants heuristic-based intuitions, participants behave as social cognitive misers and cursory accept others' responses (i.e., regardless of their competence level; Experiment 1). When at least some of the others' answers diverge from participants' intuitions this triggers epistemic vigilance mechanisms of coherence checking (Experiment 2), which make them first pay closer attention to the competence of the source. If the other is a competent source then participants accept their answers with superficial monitoring (Experiment 2 - high competence condition). Only if (some of) the other's answers are contra intuitive and the other is perceived as low in competence (Experiment 2 - low competence condition) do participants seem to finally engage on deeper monitoring of the answers and start discriminating between valid answers (accepting them) and invalid answers (rejecting and replacing them by valid ones).

If participants in the low-competence condition of Experiment 2 thought more deeply about other's answers when responding to the problems during the first block of the experiments then they should perform better (when compared to participants in the high ompetence condition) in the second block of problems (during which other participant's answers were no longer presented). That was not the case. Participants' performance in the second block of problems did not differ across the experimental conditions. However, this was not due to an absence of improvement from one block to the other, but because in both conditions participants' performance increased from the first to the second block of problems. Such overall increase in judgment performance is congruent with the view that participants behave as social cognitive misers: following other's responses if a) they are available; and b) if they agree with one's intuitions and/or if the other is a competent source of information. When there are no responses available from others (as is the case in the second block of problems) participants engage in deeper processing to solve the problems on their own. In other words, participants' performance in the first phase might have been hindered by uncritically following others' answers, which makes it difficult to discriminate between an increase in performance in the second block that results from the mere absence of other's responses (Experiment 1 and Experiment 2 - high competence condition) from an increase in performance that results from deeper processing triggered during the first block (Experiment 2 - low competence condition). Future research should discriminate between these two possible accounts.

These results contrast with those of Mata, Fiedler et al.'s (2013) experiments on the impact of being vigilant towards others' reasoning. Mata et al. showed that the more people exhibited a disposition to be vigilant towards others' answers, the more they detected and corrected the flaws of others' heuristic answers. In their experiments, the fact that other participants' answers were the heuristic and highly intuitive answers that most people give to this kind of tasks did not interfere with people's ability to detect and correct others' errors.

However, such contradiction in results may be more apparent than real. A crucial difference between Mata, Fiedler et al.'s and the present experiments is that in the first case vigilance towards others' reasoning was inferred from people's disposition to believe that others are more prone to biases than the self (i.e., BBS), while in the latter it was situationally induced by the experimental manipulation (high vs. low competence conditions). Experiment 1's findings suggest that these two types of vigilance might not be equivalent as it was initially thought. While people's disposition to be vigilant towards others' reasoning might be associated with more deeply engrained habits of scrutinizing the information provided by others, regardless of its source or its coherence with one's previous knowledge and beliefs, situationally induced vigilance might be more context dependent, thus varying in accordance to specific characteristics of the information provided, the source of the information, and the goals of the decision-maker.

In fact, these results are consistent with both the conflict elaboration theory of social influence (Butera & Mugny, 1992, 2001; Butera et al., 1995, 1996; Mugny et al., 1995; Pérez

& Mugny, 1993, 1996) and the epistemic vigilance approach (Mercier & Landemore, 2012; Mercier & Sperber, 2011, 2017; Sperber et al., 2010). It follows from both perspectives, that the extent to which a person attend to the information provided by others may change her decision space either by providing a response short-cut (i.e., imitate de other) or by leading to a reevaluation of the problem (and the problem features) after knowing others' responses. More specifically, according to these theories, considering others' responses depends both on the source of this information (e.g., high vs. low status sources) and on its (dis)similarity with people's own knowledge and beliefs. As in the present experiments, when the information provided is coherent with people's prior knowledge and beliefs, it is more likely that they accept it without further scrutiny (i.e., imitation). On the contrary, information incoherent with prior knowledge and beliefs is more easily accepted when provided by a high-status source, and more likely to be scrutinized otherwise (i.e., reevaluation of the problem).

According to the conflict elaboration theory, the aforementioned predictions specifically hold for aptitude tasks similar to those used in the present experiments, in which people know that an objective correct answer exists and that their answers are perceived as diagnostic of their aptitudes/capacities. Other types of tasks might induce different processes of solving the conflict that emerges when people are presented with information that diverges form their own knowledge and beliefs (Mugny et al., 1995; Pérez & Mugny, 1993, 1996). For instance, opinion tasks similar to those used in most persuasion research are characterized by not having an objective correct answer and by allowing others to make inferences from people's answers (i.e., specific answers/opinions assign people to specific groups). In such tasks, the conflict elaboration is shaped by the concern to maintain the categorical differentiation (e.g., ingroup agreement and outgroup disagreement) and to avoid the self-attribution of negative attributes. The most relevant characteristic of the source is not anymore its competence or benevolence (as in aptitude tasks), but its social background (i.e., being from the same or other social group or category). Therefore, although the findings from the present experiments are consistent with research on attitudes and persuasion, they are better viewed as an extension of this research to the abovementioned aptitude tasks often used in the JDM research (see also Blankenship et al., 2008; Wegener et al., 2001, 2010).

At this point, I should also mention the similarities between the present findings and those of advice-taking research. In fact, previous research on advice-taking has shown that the similarity between the advisor's information and the advisee's prior knowledge and beliefs is a strong predictor of advice acceptance (Yaniv, 2004; Yaniv & Milyavsky, 2007), along with the perceived competence and trustworthiness of the advisor (Feng & MacGeorge, 2006; Jungermann & Fischer, 2005; Van Swol & Sniezek, 2005), and the nature of the task at hand (Harvey & Fischer, 1997; Müller-Trede, Choshen-Hillel, Barneron, & Yaniv, 2017; Van Swol, 2011; Yaniv, Choshen-Hillel, & Milyavsky, 2011).

Furthering the aforementioned research on social influence, persuasion and advice-taking, the present experiments have also explored the metacognitive consequences of being exposed to others' (dis)similar answers before giving their own answers. In this regard, results from both experiments showed that when the answer provided by others was the heuristic and highly intuitive answer that most people give to these tasks, participants were less confident about their correct (rule-based) than incorrect (heuristic) answers, when answering in the high but not in the low competence condition. In the latter case, no differences were found between participants' confidence on correct and incorrect answers. When the answers provided by others were the rule-based and rarer ones (Experiment 2), participants' confidence was higher for their correct than incorrect answers, in both competence conditions. In other words, the relation between one's own answers and those provided by others strongly influenced participants' confidence, which tended to increase when both answers converged and to decrease when they conflicted.

Such results contrast with most previous research showing that people are more confident on their correct than incorrect answers (e.g., De Neys et al., 2011; Mata, Ferreira et al., 2013; Thompson, 2009; but see Johnson-Laird & Savary, 1999). These higher levels of confidence on correct answers have been interpreted as revealing a metacognitive advantage of correct responders resultant from knowing that their answer is the right one, in spite of knowing that an alternative, highly appealing but wrong answer (heuristic-based) exists (Mata, Ferreira et al., 2013; see also Mata, 2019). However, the present results seem to suggest that the metacognitive advantage that correct responders have when answering alone is weakened when, in social contexts, they are presented with others' answers that diverge from their own. Further research and theoretical development is probably needed to clarify whether and how previous metacognitive accounts of why subjective feelings of confidence vary in responding correctly or incorrectly to conflict reasoning problems can accommodate the current findings.

Participants' confidence on correct and incorrect answers has also been used to test whether people detect the conflict between heuristic answers and the logical or statistical principles evoked in the problems typically used in this field (e.g., De Neys, 2014; De Neys et al., 2011). Specifically, the fact that people are less confident when giving incorrect answers to conflict problems than when giving the correct ones to no-conflict problems (in which heuristics and logical or statistical rules converge to the same answer) has been interpreted as a sign that people implicitly detect the abovementioned conflict, even if they cannot provide the correct answer.

The present experiments provide new evidence suggesting that confidence measures might not be a reliable measure of conflict detection when other sources of potential conflict are present in the decision context. In fact, when answering in social contexts, besides the intrapsychic conflict between the heuristic and rule-based answers that is often inherent to these tasks, there is a potential (inter-psychic) conflict between the answers provided by others and one's own answers. Such inter-psychic conflict seems to dominate the intra-psychic conflict that has been the focus of attention on research stemming from dual process theories of judgment and reasoning (e.g., Evans & Stanovich, 2013, Stanovich, 2011; Thompson et al., 2013). More experiments should be conducted to better understand the interplay between these two types of conflict and whether and how it affects people's reasoning and metacognition (e.g., whether people modify their answers as consequence of experience the aforementioned inter-psychic conflict).

Remarkably, in Experiment 1, the metacognitive cost of dissenting from the heuristic and highly intuitive appealing answers provided by a highly competent other spillover to the second set of problems. Although problems in this second set were no longer presented with other's answers, participants in the high competence condition remained less confident when giving the rule-based and correct answers than when giving the heuristic-based and incorrect ones. Given that many of our judgments and decisions are made in social contexts, in which our decision processes are influenced by others from whom we receive information but also by others to whom our judgments and decisions will be communicated (see Lerner & Tetlock, 1999; Simonson & Nye, 1992 for some results on the impact of accountability), future experiments should further explore whether the decrease in confidence experienced after dissenting from other participant's answers would affect the reasoning strategies that participants use when answering to the same other participant in a new block of problems. It can be the case that being exposed to others' dissimilar answers only affects participants' metacognitive experiences (lowering the confidence on correct but dissimilar answers), but these metacognitive experiences might also impact the reasoning strategies adopted in subsequent tasks (for instance, making participants regress to the heuristic-based answers consistent with those previously provided by the other participant).

Although insightful, some of the reported results were only marginally significant, in which cases they should be interpreted with caution. Other potential weakness of these experiments, particularly in Experiment 1, is the strength of the experimental manipulation of the other participant's competence. Although at an aggregate level, the two experimental conditions differed regarding the other participant's competence, participants have often considered the stereotypical low competence group as competent and, less frequently, the stereotypical high competence group as low in competence.

To address this limitation, future experiments could let participants learn about other participant's competence through direct experience, instead of using groups stereotypically associated with low and high competence. Specifically, participants could be presented with a first block of trials in which they were presented with several problems along with the answers given by the other participant and the correspondent feedback. Participants' perceptions of other's competence would thus depend on the feedback received and should affect the extent to which participants will accept or reject other's answers in a second block of problems (where feedback is no longer provided). In the simplest version of this paradigm, the feedback provided allow people to correctly infer the competence of other's participant, and the predictions of the present experiments would hold. Extensions of this paradigm would include manipulations of the decision environment similar to those used in Chapter 2. For instance, depending on the decision environment, 1) the feedback received on other's answers may sometimes be useless to learn about her competence (e.g., giving a correct answer in a problem in which heuristic- and rule-based answers coincide is not diagnostic of one's competence), and 2) the consequences of uncritically following other's answers might be more or less pronounced (e.g., following other's answers might benefit participants' performance when other's reasoning strategies are accurate and the decision environment allows to keep using the same strategy across trials, but it could hinder participants' performance when decision environments require the use of different strategies across trials).

Moreover, to test for the impact of others' answers in one's reasoning processes, future experiments could also include problems in which the other participant gives an incorrect but not highly intuitive appealing answer. In contrast with the heuristic-based answers used in the present experiments, these incorrect answers will more easily diverge from participants' own answers and should thus facilitate the triggering of more elaborate processes when the answer comes from a low competence source. The same answers provided by a high competence source should be uncritically followed.

Likewise, future research could also explore whether characteristics of the source, other than its competence, have different impact on the extent to which people adopt a critical mindset towards the provided answers. It could be the case that when people suspect of others' motivations, they further scrutinize their answers, even when they are highly intuitive and appealing.

Regardless of all these different research avenues yet to be explored the current findings already show that reasoning and judgment embedded in social contexts are likely to be qualitatively different from those operating in the decontextualized one-shot experiments more often found in the HB tradition. In fact, under certain circumstances, social contexts might trigger the correction of erroneous judgments, as suggested by the results of Experiment 2 (Chapter 4), as well as by the two experiments presented in Chapter 3. Unfortunately, given that the trigger to engage in more elaborate reasoning processes seems to depend on the divergence between people's own answers and the answers provided by another, the debiasing effectiveness of social contexts might be reduced in the classical HB tasks, to the extent that most of us tend to give the same heuristic and highly intuitive appealing answers. In fact, this might be one more reason why the errors that result from using heuristics are so difficult to avoid. In a sense, social contexts in which other people provide the same highly intuitive appealing but incorrect answers might be seen as a particular form of the wicked environments explored in Chapter 2.

Implications for Behavioral Economics

Despite many criticisms, the HB research program has stimulated continuous and intensive research since its onset and is still a productive approach to study human judgment (Baron, 2014). One of the major implications of this research program was the emergence of behavioral economics, a discipline in the intersection of psychology and economics that uses the insights from this research program in particular (and from social psychology in general) to design behavioral interventions in diverse applied contexts such law (e.g., Alemanno & Sibony, 2015; Jolls & Sunstein, 2004; Jolls, Sunstein, & Thaler, 1998; Korobkin, 2004), economics (e.g., Benartzi & Thaler, 2013; Shefrin & Thaler, 1992), health (e.g., Downs, Loewenstein, & Wisdom, 2009; Loewenstein, Asch, & Volpp, 2013; Thorgeirsson & Kawachi, 2013), and education (e.g., Levitt, List, Neckermann, & Sadoff, 2016; O'Reilly, Chande, Groot, Sanders, & Soon, 2017). The rationale for these proposals is that people's decisions fall short of the ones proposed by rational decision-making models described in classical economic theory. It follows that psychological insights about how people actually make decisions should be incorporated into economic models and people should be helped to make better decisions (Thaler & Sunstein, 2008).

In response to these proposals classical economic theorists have claimed that the biases found in HB literature are for the most part artifactual and should disappear in contexts that better match the characteristics of real-world decisions environments, that is to say, when people answer in continuous contexts that allow them to have feedback, receive incentives for their accuracy, and interact with other social agents in the decision making contexts (e.g., Levitt & List, 2007; List & Millimet, 2008; Loomes, Starmer, & Sugden, 2003).

Earlier research already suggested that such claim might be flawed. On one hand, some previous findings have already shown that judgmental biases might persist even when incentives are provided (Camerer & Hogarth, 1999; Epley & Gilovich, 2005; Grether & Plott, 1979; Wilson et al., 1996 but see Meub et al., 2013; Simmons et al., 2010 for alternative findings), and judgments are made in social contexts (Meub & Proeger, 2015, 2018; see also Butera et al. 1996; 2001; Mata, Fiedler et al., 2013 for alternative findings). On the other hand, it is not clear that real-world decision environments are better characterized by the existence of helpful feedback – in fact, many judgmental biases have been justified as a consequence of habits acquired in environments with missing or misleading feedback (i.e., wicked environments) (Einhorn, 1980; Einhorn & Hogarth, 1978; Hogarth, 2001).

To deal with these apparently ubiquitous biases, behavioral economists developed a theory focused on changing the choice architecture of decision environments to take advantage of people's systematic errors (Thaler & Sunstein, 2008). For instance, given that people are influenced by the information that is more salient in the decision environment, the most relevant information for a given decision should be presented in a clear-cut format (e.g., (Bertrand & Morse, 2009; Salisbury, 2014). Although this approach has achieved considerable success in improving the quality of people's decision-making processes (BIT, 2019; OECD, 2017), from the perspective of the current work, these interventions might be transforming our decision environments into a particular form of wicked decision environments. In fact, by aligning decision environments with people's intuitive decision making strategies, these interventions improve the outcomes of people's decisions in the short-run but hinder their opportunities for learning about the eventual inadequacy of their intuitions. In so doing, these interventions might be leaving people more prone to be exploited by sneaky agents who often change or structure the decision environments to better serve their interests (see Sunstein, 2018; Thaler, 2018 on sludge).

Alternatives to nudge interventions have been proposed according to which people would do better decisions if the correct representation of the problems was facilitated (e.g., if natural frequencies were used instead probabilities) and if they were educated to increase their statistical literacy and decision skills (e.g., Gigerenzer, 2014; Grüne-Yanoff & Hertwig, 2016; Hertwig, 2017; Hertwig & Grüne-Yanoff, 2017). These so-called 'boost interventions' undoubtedly help people to make better decisions (e.g., Berkowitz et al., 2015; Hoffrage et al., 2000; Lindsey, Hertwig, & Gigerenzer, 2002). However, the fact that experts are far from being immune to judgmental biases (remember that many of Tversky and Kahneman participants were highly sophisticated in statistics; e.g., Tversky & Kahneman, 1971) suggests that in many occasions the cause of bad decisions is not the lack of adequate knowledge but has motivational basis. For instance, research on people's naïve theories of biases (e.g., Pronin, 2007; Pronin et al., 2002; Wilson & Brekke, 1994) showed that most people are not aware of their own susceptibility to biases, which might hinder its correction (even when people have appropriate knowledge).

By studying how classic heuristics and biases unfold in different environments, the current work contributes to this debate. Biases might be corrected in kind environments, where complete and timely feedback help people to recognize the inadequacy of their answers or strategies (e.g., Chapter 2) but they might also be exacerbated in wicked environments, where misleading feedback keep people unwarrantedly convinced of the appropriateness of their judgmental strategies (e.g., Chapter 4).

It is thus important to qualify classical economic theorists' claims. Judgmental biases found with one-shot experiments commonly used in the HB tradition might, in fact, be attenuated in some continuous and socially enriched contexts (kind environments), but they may also be amplified in (equally continuous and socially rich) wicked environments. Unfortunately, the characteristics of our real-world decision environments (e.g., frequently missing or misleading feedback), along with some aspects of people's judgment and decision-making strategies (e.g., meta-cognitive myopia), and the building blocks of human intuition (e.g., fluency, recognition, similarity, affect), suggest that the correction/attenuation of biases might be less frequent than implied by the perspective of traditional economists.

Final Considerations

Overall, results from the present experiments showed that is worth studying how judgmental biases found in discrete tasks unfold in continuous and socially enriched contexts. In line with the claims of many HB program opponents, results from the empirical work presented seem to suggest that judgmental biases found in discrete tasks might be reduced i) when, in continuous learning contexts, people are made aware of the inadequacy of their judgmental strategies and have opportunities to correct them (see Chapter 2), and 2) when cues from the social context lead them to second guess their answers (see Chapters 3 and 4). However, the role that decision environments play on the correction of people's erroneous judgments is not as clear cut as some critics might suggest. On one hand, many of the real-world decision environments do not provide people with the necessary conditions to recognize their errors and/or the inadequacy of their strategies (i.e., in Hogarth's words, they are wicked instead of kind); on the other hand, the fact that most people give the same heuristic-based answers to HB judgment and decision making tasks, seems to make its incorrectness go unnoticed and, therefore, not corrected. To put it differently, decision environments are not a panacea to correct all biases and, under some circumstances, they can even reinforce these biases.

In this thesis, I explored different manipulations of decision environments in several different experimental paradigms using a diversity of reasoning tasks that call for different heuristics and produced distinctive biases. Such in breadth approach contrasts to an in depth style that exhaustively explores different manipulations of the decision environments within one experimental paradigm.

The advantage of this option, I believe, was to provide a set of findings relevant to a large scope of theories and research on judgment, reasoning and decision making, paving the way for new theoretical extensions that reflect the role of continuous and socially enriched environments on how we, as social beings, reason and decide.

The disadvantage of the in breadth option was that the preliminary nature of some of the reported findings eventually raise more new questions and insightful avenues for future research than closing answers. However, revealing new questions and issues not yet fully considered by the state of the art is one of the hallmarks of experimental research and of the scientific endeavor to which I intend to contribute. In other words, this is just the end of the beginning.

References

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: Training counteracts the representativeness heuristic. *Cognitive Development*, *6*(2), 195–217. https://doi.org/10.1016/0885-2014(91)90036-D

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, *21*(4), 515–550. https://doi.org/10.1016/0010-0285(89)90017-0

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, *35*(5), 303–314. https://doi.org/10.1037/0022-3514.35.5.303

Alemanno, A., & Sibony, A.L. (Eds.). (2015). *Nudge and the law: A European perspective*. Bloomsbury Publishing.

Alter, A. L., Oppenheimer, D. M., & Epley, N. (2013). Disfluency prompts analytic thinking- But not always greater accuracy: Response to Thompson et al. (2013). *Cognition*, *128*(2), 252–255. https://doi.org/10.1016/j.cognition.2013.01.006

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Argote, L., Seabright, M. A., & Dyer, L. (1986). Individual versus group use of base-rate and individuating information. *Organizational Behavior and Human Decision Processes*, *38*(1), 65–75. https://doi.org/10.1016/0749-5978(86)90026-9

Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*(1), 1–10. https://doi.org/10.1016/j.jebo.2004.10.003

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486-498. https://doi.org/10.1037/0033-2909.110.3.486

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258-290. https://doi.org/10.1037/h0055756

Augustinova, M. (2008). Falsification cueing in collective reasoning: Example of the Wason selection task. *European Journal of Social Psychology*, *38*(5), 770–785. https://doi.org/10.1002/ejsp.532

Banks, A. P., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology*, *51*(3), 290–297. https://doi.org/10.1111/psyp.12169

Bar-Hillel, M. (1979). The role of sample size in sample evaluation. *Organizational Behavior and Human Performance*, *24*(2), 245–257. https://doi.org/10.1016/0030-5073(79)90028-x

Bar-Hillel, M. (1984). Representativeness and fallacies of probability judgment. *Acta Psychologica*, *55*(2), 91–107. https://doi.org/10.1016/0001-6918(84)90062-3

Barbey, A. K., & Sloman, S. (2007). Base-rate respect: From ecological rationality to dual processes. *The Behavioral and Brain Sciences*, *30*(3), 241–254; discussion 255-297. https://doi.org/10.1017/S0140525X07001653

Baron, J. (2014). Heuristics and biases. In E. Zamir, & D. Teichman (Eds.), *The Oxford Handbook of Behavioral Economics and the Law* (pp. 3–27). Oxford: Oxford University Press.

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes*, *100*(1), 76–95. https://doi.org/10.1016/j.obhdp.2005.10.004

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295. https://doi.org/10.1126/science.275.5304.1293

Behavioural Insights Team (BIT). (2019). The behavioural insights team annual update report 2017-2018. Retrieved from https://www.bi.team/publications/the-behavioural-insights-team-annual-report-2017-18/

Benartzi, S., & Thaler, R. H. (2013). Behavioral economics and the retirement savings crisis. *Science*, *339*(6124), 1152–1153. https://doi.org/10.1126/science.1231320

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10*(1)*, 122-142. https://doi.org/10.1006/game.1995.1027

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. *Science*, *350*(6257), 196–198. https://doi.org/10.1126/science.aac7427

Bertrand, M., & Morse, A. (2009). Information disclosure, cognitive biases and payday borrowing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1344420

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, *45*(6), 1185–1195. https://doi.org/10.1037//0022-3514.45.6.1185

Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology*, *44*(6), 1465–1476. https://doi.org/10.1016/j.jesp.2008.07.005

Bless, H., & Schwarz, N. (1999). Sufficient and necessary conditions in dual-mode models: The case of mood and information processing. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 423–440). New York: Guilford press.

Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes*, *88*(2), 719–736. https://doi.org/10.1016/s0749-5978(02)00010-9

Braga, J. N., Ferreira, M. B., Sherman, S. J., Mata, A., Jacinto, S., & Ferreira, M. (2018). What's next? Disentangling availability from representativeness using binary decision tasks. *Journal of Experimental Social Psychology*, *76*, 307–319. https://doi.org/10.1016/j.jesp.2018.03.006

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*(1–3), 223–241. https://doi.org/10.1016/0001-6918(80)90034-7

Brehmer, B. (1996). Man as a stabiliser of systems: From static snapshots of judgement processes to dynamic decision making. *Thinking & Reasoning*, *2*(2–3), 225–238. https://doi.org/10.1080/135467896394528

Brewer, N. T., Chapman, G. B., Schwartz, J. A., & Bergus, G. R. (2007). The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Medical Decision Making*, *27*(2), 203–211. https://doi.org/10.1177/0272989X06298595

Bröder, A., Glöckner, A., Betsch, T., Link, D., & Ettlin, F. (2013). Do people learn option or strategy routines in multi-attribute decisions? The answer depends on subtle factors. *Acta Psychologica*, *143*(2), 200–209. https://doi.org/10.1016/j.actpsy.2013.03.005

Butera, F., Caverni, J. P., & Rossi, S. (2005). Interaction with a high-versus low-competence influence source in inductive reasoning. *The Journal of Social Psychology*, *145*(2), 173–190. https://doi.org/10.3200/SOCP.145.2.173-190

Butera, F., Legrenzi, P., Mugny, G., & Perez, J. A. (1992). Influence sociale et raisonnement. *Bulletin de Psychologie*, *45*(405), 144–154.

Butera, F., Legrenzi, P., & Oswald, M. (1997). Is context a bias? *Swiss Journal of Psychology*, *56*, 59–61.

Butera, F., & Mugny, G. (1995). Conflict between incompetences and influence of a low - expertise source in hypothesis testing. *European Journal of Social Psychology*, *25*(4). https://doi.org/10.1002/ejsp.2420250408

Butera, F., & Mugny, G. (2001). Conflicts and social influences in hypothesis testing. In C. K. W. De Dreu & N. K. De Vries (Eds.). *Group consensus and minority influence implications for innovation* (pp. 160–192). Oxford, UK: Blackwell.

Butera, F., Mugny, G., Legrenzi, P., & Perez, J. A. (1996). Majority and minority influence, task representation and inductive reasoning. *British Journal of Social Psychology*, *35*(1), 123–136. https://doi.org/10.1111/j.2044-8309.1996.tb01087.x

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1–3), 7-42. https://doi.org/10.1023/A:1007850605129

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*(3), 182-191.

Chanthika, P. (2006). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*(4), 223–242. https://doi.org/10.1002/bdm.3960070402

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *The psychology of intuitive judgment: Heuristics and biases.* (pp. 120–138). New York: Cambridge University Press. https://doi.org/10.5465/AMR.2004.14497675

Chu, Y. P., & Chu, R. L. (1990). The subsidence of preference reversals in simplified and marketlike experimental settings: A note. *The American Economic Review*, *80*(4), 902-911.

Cohen, L. (1979). On the psychology of prediction: Whose is the fallacy? *Cognition*, *7*(4), 385–407. https://doi.org/10.1016/0010-0277(79)90023-4

Cohen, L. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*(3), 317-331. https://doi.org/10.1017/S0140525X00009092

Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, *10*(5), 496–502. https://doi.org/10.3758/bf03197653

Daniel, D. B., & Klaczynski, P. A. (2006). Developmental and individual differences in conditional reasoning: Effects of logic instructions and alternative antecedents. *Child Development*, *77*(2), 339–354. https://doi.org/10.1111/j.1467-8624.2006.00874.x

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin*, *28*(10), 1379–1387. https://doi.org/10.1177/014616702236869

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*(5), 428-433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187. https://doi.org/10.1080/13546783.2013.854725

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0015954

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: not always winning but at least taking part. *Cognition*, *113*(1), 45–61. https://doi.org/10.1016/j.cognition.2009.07.009

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299. https://doi.org/10.1016/j.cognition.2007.06.002

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*(5), 819-829. https://doi.org/10.1037/0022-3514.66.5.819

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory & Cognition*, *24*(5), 644–654. https://doi.org/10.3758/BF03201089

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, *43*(2), 111–121. https://doi.org/10.1016/0001-6918(79)90017-9

Doherty, M. E., Schiavo, M. B., Tweney, R. D., & Mynatt, C. R. (1981). The influence of feedback and diagnostic data on pseudodiagnosticity. *Bulletin of the Psychonomic Society*, *18*(4), 191–194. https://doi.org/10.3758/bf03333600

Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centrations in cognitive development. *European Journal of Social Psychology*, *9*(1), 105–108. https://doi.org/10.1002/ejsp.2420090110

Doise, W., Mugny, G., & Pérez, J. A. (1998). The social construction of knowledge, social marking and sociocognitive conflict. In U. Flick (Ed.), *The psychology of the social* (pp. 77–90). Cambridge: Cambridge University Press.

Doise, W., Mugny, G., & Perret-Clermont, A. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology*, *5*(3), 367–383. https://doi.org/10.1002/ejsp.2420050309

Dowd, K. W., Petrocelli, J. V, & Wood, M. T. (2014). Integrating information from multiple sources: A metacognitive account of self-generated and externally provided anchors. *Thinking & Reasoning*, *20*(3), 315–332. https://doi.org/10.1080/13546783.2013.811442

Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for promoting healthier food choices. *American Economic Review*, *99*(2), 159–164. https://doi.org/10.1257/aer.99.2.159

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin*, *31*(5), 680–692. https://doi.org/10.1177/0146167204271570

Einhorn, H. (1980). Learning from experience and suboptimal rules in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 268-284). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809477.020

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, *50*(3), 387-395. https://doi.org/10.1207/s15327752jpa5003_8

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*(5), 395–416. https://doi.org/10.1037//0033-295X.85.5.395

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: processes of judgment and choice behavioral decision theory processes of judgment and choice. *Annual Review of Psychology,* 32, 53-88. https://doi.org/10.1146/annurev.ps.32.020181.000413

Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology*, *31*(7), 1535–1551. https://doi.org/10.1111/j.1559-1816.2001.tb02687.x

Englich, B., Mussweiler, T., & Strack, F. (2005). The last word in court - a hidden disadvantage for the defense. *Law and Human Behavior*, *29*(6), 705–722. https://doi.org/10.1007/s10979-005-8380-7

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, *32*(2), 188–200. https://doi.org/10.1177/0146167205282152

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396. https://doi.org/10.1111/1467-9280.00372

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generatedand externally provided anchors. *Journal of Behavioral Decision Making*, *18*(6), 199–212. https://doi.org/10.1509/jmr.10.0513

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709-724. https://doi.org/10.1037//0003-066x.49.8.709

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/s0022-5371(81)90165-1

Evans, J. S. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*(6), 978-996. https://doi.org/10.1037//0033-2909.128.6.978

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning*, *13*(4), 321–339. https://doi.org/10.1080/13546780601008825

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T. (2012). Dual process theories of deductive reasoning: facts and fallacies. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 115–133). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0008

Evans, J. S. B. T. (2019). Reflections on reflection: The nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, *25*(4), 383-415. https://doi.org/10.1080/13546783.2019.1623071

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389. https://doi.org/10.1080/13546780542000005

Evans, J. S. B. T., & Dusoir, A. E. (1977). Proportionality and sample size as factors in intuitive statistical judgement. *Acta Psychologica*, *41*(2–3), 129–137. https://doi.org/10.1016/0001-6918(77)90030-0

Evans, J. S. B. T., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*(3), 197–213. https://doi.org/10.1016/s0010-0277(00)00098-6

Evans, J. S. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, *6*(3), 263–285. https://doi.org/10.1080/09541449408520148

Evans, J. S. B. T., & Over, D. E. (2013). *Rationality and reasoning*. Hove, England: Psychology Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, J. S. B. T., Venn, S., & Feeney, A. (2002). Implicit and explicit processes in a hypothesis testing task. *British Journal of Psychology*, *93*(1), 31–46. https://doi.org/10.1348/000712602162436

Feeney, A., Evans, J. S. B. T., & Clibbens, J. (1997). Probabilities, utilities and hypothesis testing. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the cognitive science society* (pp. 217–222). Mahwah, NJ: Erlbaum.

Feeney, A., Evans, J. S. B. T., & Clibbens, J. (2000). Background beliefs and evidence interpretation. *Thinking & Reasoning*, *6*(2), 97–124. https://doi.org/10.1080/135467800402811

Feeney, A., Evans, J. S. B. T., & Venn, S. (2008). Rarity, pseudodiagnosticity and Bayesian reasoning. *Thinking & Reasoning*, *14*(3), 209–230. https://doi.org/10.1080/13546780801934549

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal*, *71*(1), 67–85. https://doi.org/10.1080/10417940500503548

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*(5), 797–813. https://doi.org/10.1037/0022-3514.91.5.797

Fiedler, K., Hütter, M., Schott, M., & Kutzner, F. (2019). Metacognitive myopia and the overutilization of misleading advice. *Journal of Behavioral Decision Making*, *32*(3), 317-333. https://doi.org/10.1002/bdm.2109

Finucane, M. L., Alhakami, A. L. I., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17. https://doi.org/10.1002/(sici)1099-0771(200001/03)13:1<1::aid-bdm333>3.0.co;2-s

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288-299. https://doi.org/10.1037//0096-1523.1.3.288

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* (pp. 422–444). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511809477.032

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037//0096-1523.3.4.552

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, *23*(3), 339–359. https://doi.org/10.1016/0030-5073(79)90002-3

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology. General*, *120*(1), 34–45. https://doi.org/10.1037//0096-3445.120.1.34

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*(3), 253–292. https://doi.org/10.1016/0010-0285(86)90001-0

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

French, J. R. P., Jr., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Oxford, England: University Michigan.

Fricker, E. (1995). Telling and trusting: Reductionism and anti-reductionism in the Epistemology of Testimony. *Mind, 104*(414), 393–411. https://doi.org/10.1093/mind/104.414.393

Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *The American Economic Review*, *88*(4), 933–946. Retrieved from http://www.jstor.org/stable/117012

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology*, *2*(1), 83–115. https://doi.org/10.1080/14792779143000033

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*(3), 592–596. https://doi.org/10.1037/0033-295x.103.3.592

Gigerenzer, G. (2014). Breast cancer screening pamphlets mislead women. *British Medical Journal, 348*, g2636. https://doi.org/10.1136/bmj.g2636

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 513–525. https://doi.org/10.1037//0096-1523.14.3.513

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506-528. https://doi.org/10.1037//0033-295x.98.4.506

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 167–184). New York, NY, US: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, NY, US: Cambridge University Press.

Glöckner, A., & Englich, B. (2015). When relevance matters: Anchoring effects can be larger for relevant than for irrelevant anchors. *Social Psychology*, *46*(1), 4–12. https://doi.org/10.1027/1864-9335/a000214

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227-247. https://doi.org/10.1037//0096-3445.117.3.227

Goldstein, D. G., & Gigerenzer, G. (2008). The recognition heuristic and the less-is-more effect. In C. Plott & V. Smith (Eds.), *Handbook of Experimental Economics Results*, Volume 1 (pp. 987–992). Amsterdam: Elsevier Press.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Granberg, D., & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and Social Psychology Bulletin*, *21*(7), 711–723. https://doi.org/10.1177/0146167295217006

Granberg, D., & Dorr, N. (1998). Further exploration of two-stage decision making in the Monty Hall dilemma. *The American Journal of Psychology*, *111*(4), 561-579. https://doi.org/10.2307/1423551

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. S. Lichtenstein & P. Slovic (Eds.), *The Construction of Preference* (pp.77–94). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511618031.006

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax & semantics*, *Vol. 3:Speech acts.* New York: Academic Press.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407–420. https://doi.org/10.1111/j.2044-8295.1982.tb01823.x

Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, *26*(1–2), 149–183. https://doi.org/10.1007/s11023-015-9367-9

Gruppen, L. D., Wolf, F. M., & Billi, J. E. (1991). Information gathering and integration as sources of error in diagnostic decision making. *Medical Decision Making*, *11*(4), 233-239. https://doi.org/10.1177/0272989X9101100401

Harrison G. W. (1994). Expected utility theory and the experimentalists. In J. D. Hey (Ed.), *Experimental Economics* (pp. 43–73). Physica, Heidelberg. https://doi.org/10.1007/978-3-642-51179-0_4

Harvey, N. & Fischer, I. (1997). Taking advice accepting help, improving judgment. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133. https://doi.org/10.1006/obhd.1997.2697

Hertwig, R. (2017). When to consider boosting: Some rules for policy-makers. *Behavioural Public Policy*, *1*(2), 143–161. https://doi.org/10.1017/bpp.2016.14

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 593–600. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science*, *27*(2), 123–128. https://doi.org/10.1177/0963721417740645

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Hertwig R., & Ortmann A. (2005). The cognitive illusion controversy: A methodological debate in disguise that matters to economists. In R. Zwick & A. Rapoport (Eds.), *Experimental Business Research* (pp. 113–130). Springer, Boston, MA. https://doi.org/10.1007/0-387-24244-9_5

Higgins, E. T. (1997). Biases in social cognition: "Aboutness" as a general principle. In C. McGarty & S. A. Haslam (Eds.), *The message of social psychology: Perspectives on mind in society* (pp. 183-199). Cambridge, MA: Blackwell.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford Press. https://doi.org/10.4135/9781446247198

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*(2), 248–271. https://doi.org/10.1037/0033-2909.118.2.248

Hinsz, V. B., Tindale, R. S., Nagao, D. H., Davis, J. H., & Robertson, B. A. (1988). The influence of the accuracy of individuating information on the use of base rate information in probability judgment. *Journal of Experimental Social Psychology*, *24*(2), 127–145. https://doi.org/10.1016/0022-1031(88)90017-0

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*(5500), 2261-2262. https://doi.org/10.1126/science.290.5500.2261

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, *90*(2), 197–217. https://doi.org/10.1037//0033-2909.90.2.197

Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.

Hogarth, R. M. (2010). Intuition: A challenge for psychological research on decision making. *Psychological Inquiry*, *21*(4), 338–353. https://doi.org/10.1080/1047840X.2010.520260

Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 734–752. https://doi.org/10.1037/0278-7393.17.4.734

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science*, *24*(5), 379–385. https://doi.org/10.1177/0963721415591878

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology. General*, *140*(3), 434–463. https://doi.org/10.1037/a0023265

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change.* New Haven, CT, US: Yale University Press.

Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of Personality and Social Psychology*, *55*(5), 710-717. https://doi.org/10.1037/0022-3514.55.5.710

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166. https://doi.org/10.1037/e722982011-058

Jepson, D., Krantz, D. H., & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences*, *6*(3), 494-501. https://doi.org/10.1017/S0140525X00017180

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*(3), 191–229. https://doi.org/10.1016/s0010-0277(99)00015-3

Jolls, C., & Sunstein, C. R. (2004). Debiasing through Law. *The Journal of Legal Studies, 35*(1), 199-242. https://doi.org/10.2139/ssrn.590929

Jolls, C., Sunstein, C. R., & Thaler, R. (1998). A behavioral approach to law and economics. *Stanford Law Review*, *50*(5), 1471–1550. https://doi.org/10.2307/1229304

Jungermann, H., & Fischer, K. (2005). Using expertise and experience for giving and taking advice. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 157–173). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781410611826

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review*, *107*(2), 384-396. https://doi.org/10.1037/0033-295x.107.2.384

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697-720. https://doi.org/10.1037/0003-066x.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY, US: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty:*

Heuristics and biases. (pp. 414–421). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511809477.031

Kahneman, D., & Tversky, A. (1982a). On the study of statistical intuitions. *Cognition*, *11*(2), 123–141. https://doi.org/10.1016/0010-0277(82)90022-1

Kahneman, D., & Tversky, A. (1982b). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* (pp. 201–208). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.015

Kareev, Y., Halberstadt, N., & Shafir, D. (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *The Quarterly Journal of Experimental Psychology*, *46*(4), 729–742. https://doi.org/10.1080/14640749308401036

Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. H. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp 89–109). Oxford: Blackwell. https://doi.org/10.1002/9780470752937.ch5

Kern, L., & Doherty, M. E. (1982). "Pseudodiagnosticity" in an idealized medical problem-solving environment. *Journal of Medical Education*, *57*(2), 100–104. https://doi.org/10.1097/00001888-198202000-00004

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852-884. https://doi.org/10.1037/0033-295x.107.4.852

Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation: Vol. 32. Decision making from* *a cognitive perspective* (pp. 365–418). New York: Academic Press. https://doi.org/10.1016/s0079-7421(08)60315-1

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. *Cognition*, *49*(1–2), 97–122. https://doi.org/10.1016/0010-0277(93)90037-V

Klayman, J., & Ha, Y. (1987). Confirmation, discontinuation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228. https://doi.org/10.1037//0033-295x.94.2.211

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, *31*(6), 680–702. https://doi.org/10.1287/mnsc.31.6.680

Koehler, D. J., & Harvey, N. (2008). *Blackwell handbook of judgment and decision making*. Oxford: Blackwell. https://doi.org/10.1002/9780470752937

Korobkin, R. B. (2004). The problems with heuristics for law. In C. Engel, & G. Gigerenzer, (Eds.), *Heuristics and the Law* (pp. 45-60). Cambridge: The MIT Press. https://doi.org/10.7551/mitpress/3488.003.0006

Krueger, J. I. (2012, April 07). Anchoring base rates [Blog post]. Retrieved from https://www.psychologytoday.com/us/blog/one-among-many/201204/anchoring-base-rates

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*(3), 313–376. https://doi.org/10.1111/1529-1006.01431

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology,* 77(2), 221-232. https://doi.org/10.1037/0022-3514.77.2.221

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge*: *Cognitive and motivational bases.* New York: Plenum. https://doi.org/10.1007/978-1-4899-0924-4

Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., … Mannetti, L. (2005). Says who? Epistemic authority effects in social judgment. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 346 –392). San Diego, CA: Academic Press. https://doi.org/10.1016/s0065-2601(05)37006-7

Langeborg, L., & Eriksson, M. (2016). Anchoring in numeric judgments of visual stimuli. *Frontiers in Psychology*, *7*, 1–7. https://doi.org/10.3389/fpsyg.2016.00225

Larrick, R. P. (2016). The social context of decisions. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 441–467. https://doi.org/10.1146/annurev-orgpsych-041015-062445

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology*, *22*(3), 177–189. https://doi.org/10.1016/0022-1031(86)90022-3

Legrenzi, P., Butera, F., Mugny, G., & Pérez, J. (1991). Majority and minority influence in inductive reasoning: A preliminary study. *European Journal of Social Psychology, 21*(4), 359–363. https://doi.org/10.1002/ejsp.2420210408

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition*, *49*(1–2), 37–66. https://doi.org/10.1016/0010-0277(93)90035-t

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. http://dx.doi.org/10.1037/0033-2909.125.2.255

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153–174. https://10.1257/jep.21.2.153

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, *8*(4), 183–219. https://doi.org/10.1257/pol.20130358

Lichtenstein, S., Earle, T. C., & Slovic, P. (1975). Cue utilization in a numerical prediction task. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(1), 77–85. https://doi.org/10.1037/0096-1523.1.1.77

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46-55. https://doi.org/ /10.1037/h0031207

Liebert, R. M., Smith, W. P., Hill, J. H., & Keiffer, M. (1968). The effects of information and magnitude of initial offer on interpersonal negotiation. *Journal of Experimental Social Psychology*, *4*(4), 431–441. https://doi.org/10.1016/0022-1031(68)90068-1

Lindeman, S. T., van den Brink, W. P., & Hoogstraten, J. (1988). Effect of feedback on base-rate utilization. *Perceptual and Motor Skills*, *67*(2), 343–350. https://doi.org/10.2466/pms.1988.67.2.343

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2002). Communicating statistical DNA evidence. *Jurimetrics*, *43*, 147-163.

List, J. A., & Millimet, D. L. (2008). The market: Catalyst for rationality and filter of irrationality. *The BE Journal of Economic Analysis & Policy*, *8*(1), 1-53. https://doi.org/10.2202/1935-1682.2115

Loewenstein, G., Asch, D. A., & Volpp, K. G. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs*, *32*(7), 1244–1250. https://doi.org/10.1377/hlthaff.2012.1163

Loomes, G., Starmer, C., & Sugden, R. (2003). Do anomalies disappear in repeated markets? *The Economic Journal*, *113*(486), C153–C166. https://doi.org/10.1111/1468-0297.00108

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica*, *64*, 167–185. https://doi.org/10.21236/ada120844

Loschelder, D. D., Trötschel, R., Swaab, R. I., Friese, M., & Galinsky, A. D. (2016). The information-anchoring model of first offers: When moving first helps versus hurts negotiators. *Journal of Applied Psychology*, *101*(7), 995-1012. https://doi.org/10.1037/apl0000096

Maciejovsky, B., & Budescu, D. V. (2007). Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions.

Journal of Personality and Social Psychology, *92*(5), 854-870. https://doi.org/10.1037/0022-3514.92.5.854

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276-298. https://doi.org/10.1037/h0076778

Maggi, J., Butera, F., Legrenzi, P., & Mugny, G. (1998). Relevance of information and social influence in the pseudodiagnosticity bias. *Swiss Journal of Psychology*, *57*(3), 188–199.

Maggi, J., Butera, F., & Mugny, M. (1996). Conflict of incompetences: Direct and indirect influences on representation of the centimetre. *Revue Internationale de Psychologie Sociale*, *1*, 91-105.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, *38*(2), 231–248. https://doi.org/10.1037/0022-3514.38.2.231

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*(1), 11–17. https://doi.org/10.3758/BF03199552

Mata, A. (2016). Proportion dominance in valuing lives: The role of deliberative thinking. *Judgment and Decision Making*, *11*(5), 441–448.

Mata, A. (2019). Further tests of the metacognitive advantage model: Counterfactuals, confidence and affect. *Psihologijske Teme*, *28*(1), 115–124. https://doi.org/10.31820/pt.28.1.6

Mata, A., & Almeida, T. (2014). Using metacognitive cues to infer others' thinking. *Judgment & Decision Making*, *9*(4), 349-359.

Mata, A., Ferreira, M. B., & Reis, J. (2013). A process-dissociation analysis of semantic illusions. *Acta Psychologica*, *144*(2), 433–443. https://doi.org/10.1016/j.actpsy.2013.08.001

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373. https://doi.org/10.1037/a0033640

Mata, A., Fiedler, K., Ferreira, M. B., & Almeida, T. (2013). Reasoning about others' reasoning. *Journal of Experimental Social Psychology*, *49*(3), 486–491. https://doi.org/10.1016/j.jesp.2013.01.010

Mayo, R. (2015). Cognition is a matter of trust: Distrust tunes cognitive processes. *European Review of Social Psychology*, *26*(1), 283–327. https://doi.org/10.1080/10463283.2015.1117249

Mayo, R., Alfasi, D., & Schwarz, N. (2014). Distrust and the positive test heuristic: Dispositional and situated social distrust improves performance on the Wason Rule Discovery Task. *Journal of Experimental Psychology: General*, *143*(3), 985–990. https://doi.org/10.1037/a0035127

McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and bayesian inference. *Cognitive Psychology*, *26*(3), 209-239. https://doi.org/10.1006/cogp.1994.1007

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN, US: University of Minnesota Press. https://doi.org/10.1037/11281-000

Mercier, H. (2014). The function of reasoning: Argumentative and pragmatic alternatives. *Thinking & Reasoning*, *19*(3-4), 488-494. https://doi.org/10.1080/13546783.2013.819036

Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, *33*(2), 243-258. https://doi.org/10.1111/j.1467-9221.2012.00873.x

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Meub, L., & Proeger, T. (2015). Anchoring in social context. *Journal of Behavioral and Experimental Economics*, *55*, 29–39. https://doi.org/10.1016/j.socec.2015.01.004

Meub, L., & Proeger, T. (2018). Are groups 'less behavioral'? The case of anchoring. *Theory and Decision*, *85*(2), 117–150. https://doi.org/10.2139/ssrn.2397777

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85*(6), 1121-1135. https://doi.org/10.1037/0022-3514.85.6.1121

Moscovici, S. (1980). Toward a theory of conversion behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 209–239). New York: Academic Press. https://doi.org/10.1016/s0065-2601(08)60133-1

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking & Reasoning*, *4*(3), 231–248. https://doi.org/10.1080/135467898394148

Mowen, J. C., & Gentry, J. W. (1980). Investigation of the preference-reversal phenomenon in a new product introduction task. *Journal of Applied Psychology*, *65*(6), 715-722. https://doi.org/10.1037//0021-9010.65.6.715

Mugny, G., Butera, F., Sanchez-Mazas, M., & Perez, J. A. (1995). Judgments in conflict: The conflict elaboration theory of social influence.In B. Booth, R. Hirsig, A. Helminger, B. Meier, & R. Volkart (Eds.), *Perception, evaluation, interpretation* (pp. 160–168). Gottingen: Hogrefe & Huber.

Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology*, *8*(2), 181–192. https://doi.org/10.1002/ejsp.2420080204

Müller-Trede, J., Choshen-Hillel, S., Barneron, M., & Yaniv, I. (2017). The wisdom of crowds in matters of taste. *Management Science, 64*(4), 1779-1803. https://doi.org/10.1287/mnsc.2016.2660

Mussweiler, T., Englich, B., & Strack, F. (2004). Anchoring effect. In R. F. Pohl (Ed.), Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory (pp. 183-200). London: Psychology Press.

Mussweiler, T., & Strack, F. (1999). Comparing is believing: a selective accessibility model of judgmental anchoring. *European Review of Social Psychology*, *10*(1), 135–167. https://doi.org/10.1080/14792779943000044

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality & Social Psychology Bulletin*, *26*(9), 1142–1150. https://doi.org/10.1177/01461672002611010

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 759–778. https://doi.org/10.1080/14640749308401038

Nemeth, C., J. (1986). Differential contributions of majority and minority influence. Psychological Bulletin, *93*(1), 23–32. https://doi.org/10.1037/0033-295X.93.1.23

Nemeth, C. J., & Kwan, J. L. (1985). Originality of word associations as a function of majority vs. minority influence. *Social Psychology Quarterly*, *48*(3), 277–282. https://doi.org/10.2307/3033688

Nemeth, C. J., & Kwan, J. L. (1987). Minority influence, divergent thinking and detection of correct solutions. *Journal of Applied Social Psychology*, *17*(9), 788–799. https://doi.org/10.1111/j.1559-1816.1987.tb00339.x

Nemeth, C J., Mosier, K., & Chiles, C. (1992). When convergent thought improves performance: Majority versus minority influence. *Personality and Social Psychology Bulletin*, *18*(2), 139–144. https://doi.org/10.1177/0146167292182004

Nemeth, C. J., & Wachtler, J. (1983). Creative problem solving as a result of majority vs minority influence. *European Journal of Social Psychology*, *13*(1), 45–55. https://doi.org/10.1002/ejsp.2420130103

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284. https://doi.org/10.1016/0010-0277(92)90019-e

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037//1089-2680.2.2.175

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339–363. https://doi.org/10.1037//0033-295X.90.4.339

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*(1), 84–97. https://doi.org/10.1016/0749-5978(87)90046-x

O'Reilly, F., Chande, R., Groot, B., Sanders, M., & Soon, Z. (2017). *Behavioural insights for education: A practical guide for parents, teachers and school leaders.* London: Pearson.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), 147–152. https://doi.org/10.1016/j.jebo.2009.04.018

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Organisation for Economic Co-Operation and Development (OECD) (2017). *Behavioural insights and public policy: Lessons from around the world.* Paris: OECD Publishing. Retrieved from https://www.oecd.org/gov/regulatory-policy/behavioural-insights-and-public-policy-9789264270480-en.htm

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532-552. https://doi.org/10.1037/0033-295x.87.6.532

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pérez, J. A., & Mugny, G. (1993). Influences sociales: La théorie de l'élaboration du conflit. Neuchâtel, Paris

Pérez, J. A., & Mugny, G. (1996). The conflict elaboration theory of social influence. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior, Vol. 2, Small group* *processes and interpersonal relations* (pp. 191-210). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research, 11*, 668-672.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 123–205). New York: Academic Press. https://doi.org/10.1016/s0065-2601(08)60214-2

Plous, S. (1993). *The psychology of judgment and decision making.* New York, NY, England: Mcgraw-Hill Book Company.

Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology, 80*(6), 918-930. https://dx.doi.org/10.1037/0022-3514.80.6.918

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, *11*(1), 37–43. https://doi.org/10.1016/j.tics.2006.11.001

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781–799. https://doi.org/10.1037/0033-295X.111.3.781

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, *42*(4), 593.-607. https://doi.org/10.1037/0022-3514.42.4.593

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, *130*, 26–43. https://doi.org/10.1016/j.obhdp.2015.05.004

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Nww York: Appleton-Century-Crofts.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207-236. https://doi.org/10.1037/0096-3445.135.2.207

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, *32*(5), 880-892. https://doi.org/10.1037/0022-3514.32.5.880

Rossi, S., Caverni, J. P., & Girotto, V. (2001). Hypothesis testing in a rule discovery problem: When a focused procedure is effective. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 263–267. https://doi.org/10.1080/02724980042000101

Salisbury, L. C. (2014). Minimum payment warnings and information decisions. *Journal of Public Policy & Marketing*, *33*(1), 49–64. https://doi.org/10.1509/jppm.11.116

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the inability to ignore useless advice. *Experimental Psychology*, *64*(3), 170–183. https://doi.org/10.1027/1618-3169/a000361

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, *91*(6), 1080-1093. https://doi.org/10.1037/0022-3514.91.6.1080

Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General*, *111*(1), 23-59. https://doi.org/10.1037/0096-3445.111.1.23

Schwarz, N. (2014). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* New York: Psychology Press. https://doi.org/10.4324/9781315805887

Sequeira, M., Ferreira, M. B., & Almeida, T. (2013). Construção de uma bateria de problemas de julgamento e decisão em língua portuguesa. *Laboratório de Psicologia*, *11*(2), 121–142. https://doi.org/10.14417/lp.731

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, *50*(1–3), 403–430. https://doi.org/10.1016/0010-0277(94)90038-8

Shaklee, H. (1983). Human covariation judgment: Accuracy and strategy. *Learning and Motivation*, *14*(4), 433–448. https://doi.org/10.1016/0023-9690(83)90026-7

Shaklee, H., & Wasserman, E. A. (1986). Judging interevent contingencies: Being right for the wrong reasons. *Bulletin of the Psychonomic Society*, *24*(2), 91–94. https://doi.org/10.3758/bf03330513

Shefrin, H. M., & Thaler, R. H. (1992). Mental accounting, saving, and self-control. In C. Camerer, G. Loewenstein & M. Rabin (Eds.), *Advances in Behavioral Economics* (pp. 395-428). Oxford: Princeton University Press. https://doi.org/10.2307/j.ctvcm4j8j.19

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189-286). Hillsdale, NJ: LEA.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, *39*(2), 211–221. https://doi.org/10.1037/0022-3514.39.2.211

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology*, *99*(6), 917–932. https://doi.org/10.1037/a0021540

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*(3), 416–446. https://doi.org/10.1016/0749-5978(92)90020-8

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S. (2014). Two systems of reasoning: An update. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379-396). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/cbo9780511808098.024

Sloman, S., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. https://doi.org/10.1016/S0749-5978(03)00021-9

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28(1), 1-39. https://doi.org/10.1146/annurev.ps.28.020177.000245

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 780–805. https://doi.org/10.1037/a0015145

Soll, J. B., Milkman, K. L., & Payne, J. W. (2014). A user's guide to debiasing. In G. Wu & G. Keren (Eds.), *The wiley blackwell handbook of judgment and decision making* (pp. 924-951). Hoboken, USA: John Wiley & Sons. https://doi.org/10.1002/9781118468333.ch33

Soyer, E., & Hogarth, R. M. (2015). Learning from experience in nonlinear environments: Evidence from a competition scenario. *Cognitive Psychology*, *81*, 48-73. https://doi.org/10.1016/j.cogpsych.2015.08.002

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. New York: Psychology Press. https://doi.org/10.4324/9781410603432

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford: Oxford University Press.

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist*, *51*(1), 23–34. https://doi.org/10.1080/00461520.2015.1125787

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423-444. https://doi.org/10.1080/13546783.2018.1459314

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, *11*(1), 3–13. https://doi.org/10.1007/s11299-011-0093-6

Stanovich, K. E., Toplak, M. E., & West, R. F. (1998). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior*, *36,* 251–285. https://doi.org/10.1016/s0065-2407(08)00006-2

Stanovich, K. E., West, R. E., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge, MA: The MIT Press.

Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K., & Davis, J. H. (1991). Effects of assigned group consensus requirement on group problem solving and group members' learning. *Social Psychology Quarterly*, *54*(1), 25–35. https://doi.org/10.2307/2786786

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect : mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*(3), 437–446. https://doi.org/10.1037/0022-3514.73.3.437

Sunstein, C. R. (2018). Sludge and ordeals. *Duke Law Journal*. http://dx.doi.org/10.2139/ssrn.3288192

Thaler, R. H. (2018). Nudge, not sludge. *Science, 361*(6401), 431. https://doi.org/10.1126/science.aau9241

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Boston Yale University Press

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Thompson, V., & Evans, J. S. B. T. (2012). Belief bias in informal reasoning. *Thinking & Reasoning*, 18(3), 278–310. https://doi.org/10.1080/13546783.2012.670752

Thompson, V. A., Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Thorgeirsson, T., & Kawachi, I. (2013). Behavioral economics: Merging psychology and economics for lifestyle interventions. *American Journal of Preventive Medicine*, *44*(2), 185–189. https://doi.org/10.1016/j.amepre.2012.10.008

Tindale, R. S. (1993). Decision errors made by individuals and groups. In N. J. Castellano Jr (Ed.), *Individual and group decision making: Current issues* (pp. 109–124). Hillsdale: Lawrence Erlbaum Associates, Inc.

Todd, P. M., & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, *24*(2), 143–165. https://doi.org/10.1016/S0167-4870(02)00200-3

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *94*(1), 197–209. https://doi.org/10.1037/0022-0663.94.1.197

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *40*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105.https://doi.org/10.1037/h0031322

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293-315. https://doi.org/10.1037/0033-295X.90.4.293

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking & Reasoning*, *5*(2), 97–113. https://doi.org/10.1080/135467899394011

Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*(1), 35–46. https://doi.org/10.1016/0001-6918(90)90033-C

Van Swol, L. M. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, *27*(1), 103–120. https://doi.org/10.1016/j.ijforecast.2010.03.002

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology*, *44*(3), 443–461. https://doi.org/10.1348/014466604X17092

Van Wallendael, L. R. (1995). Implicit diagnosticity in an information-buying task. How do we use the information that we bring with us to a problem? *Journal of Behavioral Decision Making*, *8*(4), 245–264. https://doi.org/10.1002/bdm.3960080403

Wanzel, S. K., Schultze, T., & Schulz-Hardt, S. (2017). Disentangling the effects of advisor consensus and advice proximity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1669–1675. https://doi.org/10.1037/xlm0000396

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 509-521. https://doi.org/10.1037/0278-7393.16.3.509

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Breadth, depth, and the role of (non-) thoughtful processes in anchoring theories. *Journal of Consumer Psychology*, *20*(1), 28–32. https://doi.org/10.1016/j.jcps.2009.12.007

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, *37*(1), 62–69. https://doi.org/10.1006/jesp.2000.1431

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*(3), 506–519. https://doi.org/10.1037/a0028857

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142. https://doi.org/10.1037/0033-2909.116.1.117

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology. General*, *125*(4), 387–402. https://doi.org/10.1037/0096-3445.125.4.387

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology, 85*(3), 389-408. https://doi.org/10.1037/0022-3514.85.3.389

Wolf, F. M., Gruppen, L. D., & Billi, J. E. (1988). Use of the competing-hypoteses heuristic to reduce “pseudodiagnosticity.” *Journal of Medical Education*, *63*, 548–554. https://doi.org/10.1097/00001888-198807000-00006

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Yaniv, I. (2004). Receiving other people’s advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*, *115*(1), 111–120. https://doi.org/10.1016/j.obhdp.2010.11.006

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120. https://doi.org/10.1016/j.obhdp.2006.05.006

Appendices

Appendix A - Materials used in the experiments presented in Chapter 2

A1.1. Structure of decision environments used in Experiment 1

Kind Environment

Type 1	Disease A	Disease B
S1	**high**	high and >a
S2	high	high and > c

Type 2	Disease A	Disease B
S1	**high**	high and > a
S2	low	high

FILLER 1	Disease A	Disease B
S1	**low**	high
S2	high	low

FILLER 2	Disease A	Disease B
S1	**high**	high and > a
S2	low	low but > c

Wicked Environment

Type1	Disease A	Disease B
S1	**high**	high and < a
S2	high	high and < c

Type 2	Disease A	Disease B
S1	**high and > b**	high
S2	high	low

FILLER 1	Disease A	Disease B
S1	**high**	low
S2	low	high

FILLER 2	Disease A	Disease B
S1	**high**	high but < a
S2	low but > d	low

A1.2. List of values presented in each trial

Type 1	Disease A	Disease B	P1	Type 1	Disease A	Disease B
S1	**86%**	98%		**S1**	**86%**	28%
S2	70%	84%		**S2**	98%	46%

Type 2	Disease A	Disease B	P2	Type 2	Disease A	Disease B
S1	**72%**	83%		**S1**	**83%**	72%
S2	43%	75%		**S2**	75%	43%

FILLER 1	Disease A	Disease B	P3	FILLER 1	Disease A	Disease B
S1	**26%**	68%		**S1**	**71%**	26%
S2	71%	16%		**S2**	16%	74%

FILLER 2	Disease A	Disease B	P4	FILLER 2	Disease A	Disease B
S1	**88%**	94%		**S1**	**88%**	77%
S2	37%	44%		**S2**	40%	29%

Type 1	Disease A	Disease B	P5	Type 1	Disease A	Disease B
S1	**62%**	75%		**S1**	**68%**	44%
S2	67%	79%		**S2**	67%	39%

Type 2	Disease A	Disease B	P6	Type 2	Disease A	Disease B
S1	**65%**	70%		**S1**	**70%**	65%
S2	41%	64%		**S2**	64%	41%

FILLER 1	Disease A	Disease B	P7	FILLER 1	Disease A	Disease B
S1	**33%**	69%		**S1**	**63%**	38%
S2	73%	20%		**S2**	33%	69%

FILLER 2	Disease A	Disease B	P8	FILLER 2	Disease A	Disease B
S1	**60%**	78%		**S1**	**60%**	54%
S2	35%	51%		**S2**	49%	35%

A1.3. Instructions used in Experiment 1

Bem-vindo(a)!

Vai participar num estudo sobre tomada de decisão em condições de incerteza.

No nosso dia-a-dia somos frequentemente confrontados com situações em que temos de tomar decisões com base num conjunto limitado de informações, quer porque não dispomos do tempo necessário para procurar toda a informação relevante, quer porque esta procura implica custos em que nem sempre podemos incorrer.

Neste estudo estamos interessados em compreender melhor o modo como as pessoas selecionam informação para tomar decisões neste tipo de situações. Para tal, vamos apresentar-lhe um conjunto de cenários de decisão médica e pedir-lhe que faça o diagnóstico dos pacientes apresentados com base num conjunto limitado de informação.

Pressione a barra de espaços para continuar

###

As instruções que se seguem são essenciais para realizar o estudo. Por favor leia-as com atenção.

Em cada cenário de decisão ser-lhe-á apresentado um paciente que apresenta 2 sintomas. A sua tarefa será decidir qual das duas doenças possíveis (no exemplo, doenças A ou B) afeta o paciente.

Para tal ser-lhe-á apresentada uma tabela onde poderá consultar a prevalência dos sintomas do seu paciente em duas doenças igualmente frequentes:

	Doença A	**Doença B**
Sintoma 1	W	X
Sintoma 2	Y	Z

Em cada cenário poderá pedir informação acerca de 2 das 4 células desconhecidas.

Especificamente, poderá pedir informação acerca da:

- prevalência de um mesmo sintoma nas duas doenças (células W e X ou células Y e Z);

- prevalência dos dois sintomas apresentados numa das doenças (Células W e Y ou células X e Z);
- prevalência de um sintoma numa doença e do outro sintoma na outra doença (células W e Z ou Y e X).

Pressione a barra de espaços para continuar

###

Todos os sintomas e doenças que serão apresentados neste estudo são fictícios, e foram incluídos de forma a impedir que utilize outro conhecimento que possa ter sobre doenças e sintomas que não seja o que é apresentado nas tabelas.

Depois de receber informação sobre as 2 células selecionadas deverá fazer o diagnóstico do paciente apresentado.

Note que todas as doenças apresentadas têm igual prevalência na população e os sintomas apresentados em cada cenário variam aleatoriamente no grau de informatividade. Por exemplo, os sintomas podem ser muito frequentes numa doença e pouco noutra, mas podem também ser igualmente pouco frequentes (sintoma 1) ou igualmente frequentes (sintoma 2) em ambas as doenças.

	Doença A	**Doença B**
Sintoma 1	5%	5%
Sintoma 2	99%	99%

Para se familiarizar com o tipo de problemas que irá resolver, apresentamos-lhe de seguida um ensaio de treino.

Pressione a barra de espaços para continuar

###

Imagine que é médico(a) e é enviado(a) em missão humanitária para uma ilha remota do Oceano Pacífico.

Chegado(a) à ilha, um conjunto de habitantes locais procuram a sua ajuda para o diagnóstico e tratamento da doença que os afeta.

Cada paciente apresenta 2 sintomas potencialmente associados a doenças características daquela região.

Para diagnosticar a doença dos seus pacientes poderá comparar os sintomas que estes apresentam com o que é conhecido acerca desses sintomas nas doenças apresentadas.

Pressione a barra de espaços para continuar

###

Uma vez que não está muito familiarizado(a) com as doenças da região, existe na ilha um especialista de cada uma das doenças a quem poderá recorrer para obter informação acerca da prevalência dos sintomas apresentados em cada uma das doenças. Uma vez que estes especialistas se encontram bastante atarefados, poderá apenas fazer 2 pedidos de informação (os dois pedidos a um mesmo especialista ou um pedido a cada um deles).

Com base nas informações prestadas pelo(s) especialista(s), pedimos-lhe faça um diagnóstico provisório da doença que lhe parece estar a afetar o seu paciente.

Para facilitar diagnósticos futuros, os seus pedidos são registados e toda a informação existente na ilha acerca da prevalência dos sintomas apresentados nas duas doenças será compilada e ser-lhe-á apresentada logo que esteja disponível. Nessa altura terá oportunidade de rever o seu diagnóstico, podendo mantê-lo ou alterá-lo em função da nova informação de que dispõe.

Pressione a barra de espaços para continuar

###

Paciente 1

O seu primeiro paciente apresenta descalcificação e alteração na produção de enzimas PTZ-3. Para o/a ajudar a fazer o diagnóstico do paciente pode pedir informação acerca de 2 das 4 células da tabela abaixo apresentada:

	Doença de Iazasper	**Doença de Mastla**
Descalcificação	W	X
Alteração na produção de enzimas PTZ-3	Y	Z

Que informação quer pedir para o/a ajudar fazer o diagnóstico?
Para dar a sua resposta, pressione a tecla correspondente à 1ª célula da tabela que quer conhecer (i.e., W, X, Y ou Z).

###

[in case that participant's first choice was 'W', the experiment continued as following:]

Paciente 1

O seu primeiro paciente apresenta descalcificação e alteração na produção de enzimas PTZ-3.

	Doença de Iazasper	**Doença de Mastla**
Descalcificação	75%	X
Alteração na produção de enzimas PTZ-3	Y	Z

Que outra informação escolheria pedir para fazer o seu diagnóstico?
Para dar a sua resposta, pressione a tecla correspondente à 2ª célula da tabela que quer conhecer (i.e., X, Y ou Z), seguida de ENTER.

###

[in case that participant's second choice was 'Z', the experiment continued as following:]

	Doença de Iazasper	**Doença de Mastla**
Descalcificação	75%	X
Alteração na produção de enzimas PTZ-3	Y	24%

Com base na informação de que agora dispõe, qual pensa ser a doença que afeta o paciente apresentado?

a) Doença de Iazasper (pressione a tecla A)

b) Doença de Mastla (pressione a tecla B)

###

Usando a escala abaixo indique, por favor, qual o seu grau de confiança na resposta anterior:

1 – nada confiante a 6 – totalmente confiante

###

	Doença de Iazasper	**Doença de Mastla**
Descalcificação	75%	43%
Alteração na produção de enzimas PTZ-3	86%	24%

Com base na nova informação apresentada, qual pensa ser a doença que afeta o paciente apresentado?

a) Doença de Iazasper (pressione a tecla A)

b) Doença de Mastla (pressione a tecla B)

###

Usando a escala abaixo indique, por favor, qual o seu grau de confiança na resposta anterior:

1 – nada confiante a 6 – totalmente confiante

A1.3.1. List of diseases and symptoms used in each trial (Experiments 1 and 2)

	Disease A	Disease B	Symptom 1	Symptom 2
T	Iazasper's Disease	Mastla's Disease	Decalcification	Change in PTZ-3 enzyme production
P1	Norvale's Disease	Elfhedge's Disease	Peristaltic suppression	Luterian cell microtrophy
P2	Llonegnim's Disease	Jerguntzen's Disease	Reduction of TPA blood levels	Xerostosis
P3	Pricote's Disease	Norpond's Disease	Epistaxomya	Pyramidal vision
P4	Tlemin's Disease	Crepkin's Disease	Increase in TREPS production	Reduction of WF-77 blood levels
P5	Redlea's Disease	Blaipikin's Disease	Vasculosis	Increased levels of EW-Z3
P6	Norbeach's Disease	Shadelake's Disease	Basal spheropathis	Light hyposalis
P7	Freizlpin's Disease	Ornilapzin's Disease	Protoleonesis	Mutation of gene FSD2
P8	Galhump's Disease	Umblenin's Disease	Allergic dendrionitis	Acute panfleasis

A2.1. Structure of decision environments used in Experiment 2

Kind Environment

Type 1	**Disease A**	**Disease B**
S1	**high**	high and > a
S2	high	high and > c

Type 2	**Disease A**	**Disease B**
S1	**high**	high and > a
S2	low	high

Type 3	**Disease A**	**Disease B**
S1	high	high and >a
S2	**low**	low mas <c

FILLER 1	**Disease A**	**Disease B**
S1	**low**	high
S2	high	low

FILLER 2	**Disease A**	**Disease B**
S1	low	high
S2	**high**	low

Wicked Environment

Type1	**Disease A**	**Disease B**
S1	**high**	high and < a
S2	high	high and < c

Type 2	**Disease A**	**Disease B**
S1	**high and > b**	high
S2	high	low

Type 3	**Disease A**	**Disease B**
S1	high and > b	high
S2	**low and < d**	low

FILLER 1	**Disease A**	**Disease B**
S1	**low**	high
S2	high	low

FILLER 2	**Disease A**	**Disease B**
S1	low	high
S2	**high**	low

A2.2. List of values presented in each trial

Kind Environment | **Wicked Environment**

Type1	Disease A	Disease B		Type1	Disease A	Disease B
S1	**88%**	98%	P1	S1	**88%**	70%
S2	70%	84%		S2	98%	84%

FILLER 1	Disease A	Disease B		FILLER 1	Disease A	Disease B
S1	**26%**	68%	P2	S1	**26%**	68%
S2	71%	16%		S2	71%	16%

Type 2	Disease A	Disease B		Type 2	Disease A	Disease B
S1	**72%**	83%	P3	S1	**83%**	72%
S2	43%	75%		S2	75%	43%

Type 3	Disease A	Disease B		Type 3	Disease A	Disease B
S1	88%	94%	P4	S1	94%	88%
S2	**42%**	37%		S2	**37%**	42%

FILLER 2	Disease A	Disease B		FILLER 2	Disease A	Disease B
S1	33%	69%	P5	S1	33	69
S2	**73%**	20%		S2	**73**	20

Type 2	Disease A	Disease B		Type 2	Disease A	Disease B
S1	**65%**	70%	P6	S1	**70%**	65%
S2	41%	64%		S2	64%	41%

Type 3	Disease A	Disease B		Type 3	Disease A	Disease B
S1	70%	78%	P7	S1	78%	70%
S2	**26%**	20%		S2	**20%**	26%

Type1	Disease A	Disease B		Type1	Disease A	Disease B
S1	**62%**	75%	P8	S1	**75%**	62%
S2	67%	79%		S2	79%	67%

A2.3. Instructions used in Experiment 2

A2.3.1. Learning phase

"This is a study about how people make decisions under uncertainty.
Next, you will be presented with several medical scenarios. In each one, you will be asked to make a diagnosis based on a limited amount of information. It is very important for you to follow all the instructions, please read them carefully!"

Please press >> button to start the study

###

"Imagine that you are a missionary doctor newly established on a remote island. One of the local residents has contracted a disease which you are having trouble diagnosing. Although the disease presents a number of common symptoms, you have been able to narrow the possibilities down to two diseases, Tropical Disease A and Tropical Disease B.
In order to begin the proper treatment, you must decide which disease is troubling your patient. One method you could use would be to compare the characteristics of this patient's disease with what is known about the symptoms of the two diseases. There is an expert medical consultant for each disease, one located in Ohio, the other in California. From your island you can place phone calls to their clinics and find out what percentage of people suffering from each of the diseases have the symptoms that your patient presents.
However, since both consultants are very busy, you will be able to get only partial information about the diseases.

###

You have already figured out that your first patient has the two symptoms presented on the first column of the table below. To make your diagnosis possible, you can now ask for information about the likelihood of these symptoms in either disease but you will be able to get access to just 2 of the 4 concealed cells:

	Norvale's Disease	**Elfhedge's Disease**
Peristaltic suppression	A	B
Luterian cell microtrophy	C	D

What cells do you want to uncover in order to diagnose what disease is affecting your patient? To give your answer, please press the key correspondent to the 1st cell that you want to uncover (i.e., A, B, C or D) and wait until a percentage is shown on it.
Please don’t press any other key while the system is processing your request!

###

[in case that participant’s first choice was cell A, the experiment continued as following:]

	Norvale's Disease	**Elfhedge's Disease**
Peristaltic suppression	86%	B
Luterian cell microtrophy	C	D

Which other cell do you want to uncover to make your diagnosis?
To give your answer, please press the key correspondent to the 2nd cell that you want to uncover and wait until a percentage is shown on it.

###

[in case that participant’s second choice was cell C, the experiment continued as following:]

	Norvale's Disease	Elfhedge's Disease
Peristaltic suppression	86%	B
Luterian cell microtrophy	98%	D

Based on the above-presented information, which disease do you think your patient has contracted? a) Norvale’s Disease; b) Elfhedge's Disease

###

How confident are you about your diagnosis? (1 – Not confident at all to 6 – Totally confident)

###

We will now present you with information about all the records made on the island:

	Norvale's Disease	Elfhedge's Disease
Peristaltic suppression	86%	28%
Luterian cell microtrophy	98%	46%

Based on the new information above-presented, which disease do you think your patient has contracted?

###

How confident are you about your diagnosis? (1 – Not confident at all to 6 – Totally confident)

###

<u>2.3.2. Final test</u>

Now that you already have some experience in diagnosing tropical diseases, you were invited to participate in the monthly medical missionary meetings where more complex cases are discussed.

Press key “5” to continue.

###

‘Patient Z’ was the last case presented for discussion in the meeting. Patient Z has all the symptoms presented on the table below. These symptoms were previously found in two different diseases, equally common in the island: Asplya and Tussau Diseases.

After discussing the case, you became in charge to call the expert(s) consultant(s). As in the previous scenarios, the information you can ask for is limited. In this particular case, you can make 6 questions.

Please use keys A to X to select the 6 cells that you want to uncover.

This time you will not receive the information about the chosen cells immediately after having made your choices.

	Asplya's Disease	Tussau's Disease
Cartilage thickening	A	B
Mutation of gene BTo9	C	D
High-frequency hearing loss	E	F
Presence of toxin ZWA in the blood	G	H
Increase in PTL production	I	J
Geometric vision	K	L
VQA marker changes	M	N
Temporal disorientation	O	P
Changes in the production of PIPP enzyme	Q	R
Loss of proprioception	S	T
Reduction of TRZ blood levels	U	V
Loss of protosympathetic function	W	X

Appendix B – Materials used in the experiments presented in Chapter 3

B1.1. General knowledge questions used in Experiment 1

	Low anchors	High anchors
What year did the Ford T come out?	1890	1999
What is the average length of a whale (in meters)?	5	50
When did the Pombaline government end?	1765	1889
When did Portugal abolish the death penalty?	1700	1940
How long is Portugal's coastline? (in km)	300	1500
When did Shakespeare die?	1600	1859
What is the highest speed of subway trains in Lisbon (km/h)	60	120
When was penicillin first discovered?	1850	1960
How many years married is ruby wedding?	27	80
How many floors does the Empire State Building have?	30	120
What percentage of air is oxygen?	5	80
What is the highest speed of a Formula 1 car? (km/h)	240	400
When was Romeo and Juliet first performed on stage?	1580	1920
In what year did Picasso paint “Guernica"?	1750	1945

B1.2. Instructions used in Experiment 1

All participants were presented with the same paragraph introducing the experiment [variations presented in brackets correspond to the differences resultant from the Source manipulation]:

> "This study aims to better understand how people make decisions in their daily lives, even when they are not sure about the correct answer. In other words, we are interested in studying how people make estimates under uncertainty. For that purpose, you will be presented with several general knowledge questions. Each of these questions will be followed by an answer [vs. Each of these questions will be followed by an answer. These answers were given by other people who have previously participated in this study]."

Participants in the No forewarning conditions (NFNS and NFS) were then presented with the specific instructions for the task:

> "For each question you will be asked to:
>
> - Indicate to what extent you agree with the presented answer [vs. other participant's answer];
> - Provide your own answer to the question (which may or may not coincide with the presented answer [vs. other participant's answer].
>
> Next, you will be presented with 2 training trials to make the task clearer."

While participants in the Forewarning conditions (FNS and FS) were presented with the forewarning about the anchoring effect, and asked to try to avoid it:

> “Previous research has shown that the mere suggestion of a possible response influences participants' own responses to a given question. Thus, even when participants reject the suggestion and give a different answer, it tends to be too close to the suggested response. Imagine that in a game of “True or False", the following question is posed: "Julius Caesar died at the age of 100. True or false?" The player responds “false" and hits. Now imagine another player who asks the question "Julius Caesar died at the age of 20. True or false?" The player responds "false" and also hits. Later someone asks these players "How old was Julius Caesar when he died?"
>
> First player’s answer will certainly be less than 100 years; and that of the second player more than 20 years! But the way they spontaneously came to their answers was by starting from the rejected value (100 years or 20 years) and then adjusting (downward in the first case and upward in the second). Because the adjustment is typically insufficient, the given answer turns out to be too close to the initial question. Thus the first player will tend to suggest that Julius Caesar died later than the second player (of course this effect only happens when players do not know that Julius Caesar was murdered at age 56).
>
> In this study we want to know to what extent people can counteract this tendency. So in the following questions you should NOT let the suggested answers influence your own judgment.
>
> Next, you will be presented with 2 training trials to make the task clearer."

B1.3. Illustration of the trials used in Experiment 1

When was cinema invented?

1800

Do you agree with this answer [vs. with answer given by this participant]?

Strongly disagree – Disagree – Agree – Totally agree

###############

When was cinema invented?

1800

What is your answer to the above question?

____

B1.4. Supplementary analyses of Experiment 1

Table 1

Mean values (95% CI between brackets) for agreement with the provided responses (anchors), response times (milliseconds), participants' subjective knowledge, subjective performance, and perceived influence by forewarnings about the anchoring effect and the source of the anchors.

	No Forewarning		Forewarning	
	No Source	Source	No Source	Source
Agreement	2.45 [2.21, 2.68]	2.30 [2.03, 2.57]	2.26 [2.05, 2.46]	2.10 [1.89, 2.31]
Anchoring score*	0.69 [0.55, 0.84]	0.52 [0.35, 0.68]	0.53 [0.41, 0.65]	0.42 [0.28, 0.56]
Response Times*	4133.75 [3437.95, 4829.55]	4785.56 [3909.26, 5661.86]	5031.06 [4329.72, 5732.40]	5979.49 [4798.96, 7160.02]
Subjective Knowledge	1.70 [1.39, 2.01]	1.80 [1.51, 2.09]	1.76 [1.48, 2.05]	2.05 [1.73, 2.37]
Subjective Performance	1.75 [1.45, 2.05]	1.80 [1.51, 2.09]	1.59 [1.33, 1.85]	2.00 [1.63, 2.37]
Subjective Influence	3.05 [2.73, 3.37]	2.60 [2.28, 2.92]	2.59 [2.22, 2.95]	2.70 [2.43, 2.97]

Note.* These values are presented in Figures 1 and 2.

Subjective Knowledge, Subjective Performance and Subjective Influence. Results from the 2 x 2 ANOVAs with forewarning and source as between-subjects factors revealed no differences across the experimental conditions either on participants' subjective knowledge (Fs = [0.41 – 1.18], ps = [.188 – .525]) or on participants' subjective performance (Fs = [0.02 – 2.40], ps = [.126 – .898]). Mean values of subjective knowledge (M = 1.83, 95% CI = [1.53, 2.13]) and subjective performance (M = 1.79, 95% CI = [1.48, 2.09]) were quite low and positively correlated (see Table 1), expressing the perceived difficulty of the task. Both of these variables show a negative correlation with subjective influence, suggesting that the more knowledgeable participants believed to be and the better they believed to have performed, the less was the subjective influence of the anchors. However, none of these variables correlated with actual performance (as measured by the mean anchoring score).

The subjective influence of the anchor-values differed across the experimental conditions as revealed by a marginally significant Forewarning x Source interaction, $F(1, 73)= 3.44$, $p = .068$, $\eta_p^2 =.05$. For participants in the no forewarning conditions, the subjective influence of the anchors was lower when they were presented as other participants' answers than without a specified source. On the contrary, when participants were forewarned about the anchoring effects, the subjective influence of the anchors was higher for socially driven anchors than for the same anchors presented without a specified source (see Table 1, fifth row). The subjective influence of the anchor-values (M = 2.74, 95% CI = [2.42, 3.05]) was the only variable that correlated with actual performance ($r(78)=.37$, $p = .001$) (see Table 2).

Table 2

Zero-order correlations between actual performance (anchoring score), subjective performance, subjective influence and subjective knowledge.

	Subjective Performance	Subjective Influence	Subjective Knowledge
Anchoring score	-.18	.37*	-.18
Subjective Performance		-.33*	.55*
Subjective Influence			-.44*

* p<.01

B2.1. General knowledge questions used in Experiment 2

	Low anchors	High anchors
What is the height of the tallest redwood (in feet)?	65	550
In what year was telephone invented?	1850	1920
How many countries are members of United Nations?	14	127
What is the amount of meat eaten per year by the average American (in pounds)?	50	1000
What is the height of Mount Everest (in feet)?	2000	45500
How many female professors are there at the University of California, Berkeley?	25	130
What is the distance between San Francisco and New York (in miles)?	1500	6000
How many babies are born in United States each day?	100	50000
What is the length of Mississippi river (in miles)?	70	2000
What is the population of Chicago (in millions)?	0,2	5

B2.2. Instructions used in Experiment 2

As in Experiment 1, all participants were presented with the same paragraph introducing the experiment [variations presented in brackets correspond to the differences resultant from the Source manipulation]:

> "In this study you will be presented with a set of difficult general knowledge questions. Each question will be followed by a possible answer [vs. followed by an answer from a previous participant in the study]."

Participants in the No forewarning conditions (NFNS and NFS) were then told that their task would be to give their own answers to each question, while participants in the Forewarning conditions (FNS and FS) were presented with the forewarning about the anchoring effect, and asked to try to avoid it:

> "Previous research has demonstrated that when facing different kinds of questions we are often influenced by possible responses that are provided to us. The goal of this study is to evaluate to what extent you can avoid this tendency. Specifically, to what extent can you resist being influenced by external sources of information when questioned about general knowledge themes.
>
> Next, you will be presented with a set of difficult general knowledge questions. Each question will be followed by a possible answer [vs. followed by an answer from a previous participant in the study]. Your task is to provide your OWN answer to each question without being influenced by the presented answer [vs. answer given by the previous participant]."

B2.3. Illustration of the trials used in Experiment 2

What is the height of the tallest redwood (in feet)?

Answer [vs. James' answer]: 65 feet

Please give YOUR OWN answer to this question. In case you don't know the answer, please give your best estimate.

____

##############

Please indicate how confident you are in your answer (1- Not at all confident – 9 Totally confident)

____

B2.4. Supplementary analyses of Experiment 2

Table 3

Mean (95% CI between brackets) for confidence ratings, participants' subjective knowledge, relative subjective knowledge, and subjective influence by forewarnings about the anchoring effect and the source of the anchors.

	No Forewarning		Forewarning	
	No Source (NSNF)	Source (SNF)	No Source (NSF)	Source (SF)
Anchoring	0.51 [0.39, 0.64]	0.50 [0.38, 0.63]	0.60 [0.46, 0.74]	0.35 [0.21, 0.48]
Confidence	3.62 [3.01, 4.23]	3.83 [3.15, 4.52]	4.06 [3.44, 4.68]	4.06 [3.43, 4.69]
Subjective knowledge	3.53 [2.90, 4.17]	3.37 [2.68, 4.06]	3.45 [2.79, 4.10]	3.32 [2.66, 3.98]
Relative subjective knowledge (own-other)	-0.13 [-0.65, 0.38]	-0.80 [-1.40, -0.20]	-0.59 [-1.25, 0.08]	-0.77 [-1.33, -0.22]
Subjective Influence	5.67 [4.90, 6.43]	4.83 [4.11, 5.55]	4.07 [3.26, 4.88]	3.71 [3.00, 4.42]

Note.* These values are presented in Figure 3.

Subjective Knowledge, Relative subjective knowledge, and Subjective Influence. Participants reported low levels of knowledge about the topics covered by the questions, with mean values below the scale mid-point in all the experimental conditions, all *t*s = [4.74; 5.20], *p*s < .001. A 2 x 2 ANOVA with forewarning and source as between-subjects factors showed that subjective knowledge did not differ across conditions, *F*s < 1.

The extent to which participants consider themselves better (or worse) than others might influence the degree of adjustment from others' answers. To explore this possibility, we computed a measure of participants' subjective knowledge (in comparison to others) by subtracting participants' subjective ratings of how much they were knowledgeable about the topics covered by the general knowledge questions from participants' ratings of how much they thought others were knowledgeable about these topics. Mean values on this measure are negative for all the experimental conditions (see Table 3, third row), meaning that participants perceive themselves as less knowledgeable about the topics than others. These results, together with the low reported levels of subjective knowledge, are congruent with previous findings pointing to a worse-than-average effect for difficult tasks (Kruger, 1999); see also Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). Perceiving others as more knowledgeable than the self could lead participants to rely more on others' answers. However, in line with what was found for subjective knowledge, the correlation between relative subjective knowledge and anchoring was close to zero (see Table 4). A 2 x 2 ANOVA with forewarning and source as between-subjects factors showed that relative subjective knowledge scores did not differ across conditions, all *F*s = [0.56; 2.23], *p*s = [.138; .457].

The subjective influence of the anchors was low to moderate (see Table 3, fourth row). Results from a 2 x 2 ANOVA with forewarning and source as between-subjects factors showed that the perceived influence of the anchors was lower among participants who were forewarned about the anchoring effects ($M = 3.88$, 95% CI = [3.36, 4.41]) than among those who were not forewarned ($M = 5.25$, 95% CI = [4.73, 5.77]), $F(1,116) = 13.73$, $p < .001$, $\eta p^2 = .11$. In agreement with previous research (Wilson et al., 1996), we found a significant correlation between the subjective influence of the anchors and the anchoring effect, $r(120) = .30$, $p = .001$. However, as discussed by Wilson et al. (1996), it is unclear whether this correlation represents an awareness of the anchoring process or a post hoc inference about how the anchor must have influenced participants.

Table 4.

Zero-order correlations between performance (anchoring score), confidence, subjective influence, subjective knowledge, and subjective relative knowledge.

	Subjective Influence	Subjective Knowledge	Subjective relative Knowledge
Anchoring	.30*	.06	.06
Confidence	.05	.73*	.36**
Subjective Influence		.09	.05

* p < .01

** p < .001

Appendix C - Materials used in the experiments presented in Chapter 4

C1.1. Reasoning problems used in Experiment 1

<u>Syllogisms</u>

S1. Believable Invalid:
All trees have branches.
A pine has branches.
Therefore, a pine is a tree.

S2. Believable Invalid:
All stars shine.
The Sun shines.
Therefore, the Sun is a star.

S2. Believable Valid (No-conflict version):
All stars shine.
The Sun is a star.
Therefore, the Sun shines.

S3. Unbelievable Valid:
All fruits have vitamins.
Caramel is a fruit.
Therefore, caramel has vitamins.

S4. Unbelievable Valid:
All animals have eyes.
Viruses are animals.
Therefore, viruses have eyes.

S4. Unbelievable Valid:

All birds have wings.

Cats are birds.

Therefore, cats have wings

Semantic Illusions

SI1. The goat is a farm animal that provides milk, cheese and wool.

SI2. It was Pythagoras who established the theorem about the relation between the sides of a rectangle.

SI3. Lava is the geologic material that a volcano expels during an earthquake.

SI4. Charles Darwin changed Biology forever when he formulated the rules of modern genetics.

SI5. The court is the place where the witnesses are judged.

SI5. No-conflict version: The court is the place where the defendants are judged.

Transitive Reasoning

TR1. Unbelievable Valid:

July is hotter than August.

Twag is hotter than August.

December is hotter than Twag.

December is hotter than August.

TR2. Unbelievable Valid:

Giraffes are bigger than elephants.

Zoots are bigger than elephants.

Mice are bigger than zoots.

Mice are bigger than elephants.

TR3. Believable Invalid:

Days are longer than erps.

Erps are longer than seconds.

Minutes are longer than seconds.

Days are longer than minutes.

TR3. Believable Valid (No-conflict version):

Days are longer than erps.

Erps are longer than minutes.

Weeks are longer than minutes.

Days are longer than minutes.

TR4. Believable Invalid:

Leopards are faster than quigs.

Quigs are faster than slugs.

Snails are faster than slugs.

Leopards are faster than snails.

TR5. Believable Invalid:

Toffees are sweeter than chocolates.

Toffees are sweeter than hoke.

Hoke are sweeter than potatoes.

Chocolates are sweeter than potatoes.

<u>Proportion Dominance</u>

PD1. Imagine that in a town, two houses are on fire: a big house with 100 people inside and a small house with 10 people inside. In this town there is only one fire engine, so you have to decide which of the two houses you will send the fire engine to. You only have time to save one of the houses. If you choose to send the fire engine to the small house, you are certain to save exactly 10 out of the 10 people in that house. If you choose to send the fire engine to the big house, you are certain to save exactly 11 out of the 100 people in that house. What house would you send the fire engine to?
(Correct answer: Big house; Intuitive answer: Small house).

PD2. An amusement park is nearing the final stages of planning before construction when it is found that construction will destroy some trees where an endangered species of songbird nests. The planners are willing to adopt one of two proposed solutions to the problem. Program A saves 19 of the 25 birds that nest in Area A. Program B saves 20 of the 400 birds

that nest in Area B. These programs are mutually exclusive and the only two options available. Which would you choose?
(Correct answer: Program B; Intuitive answer: Program A).

PD3. The current recession has forced companies to cut jobs. Your office provides financial support to struggling businesses in the local economy, but limited resources force you to choose which businesses to assist. Program A saves 56 of the 560 jobs that would have other-wise been lost at Factory A. Program B saves 54 of the 60 jobs that would have otherwise been lost at Factory B. These programs are mutually exclusive and the only two options available. Which would you choose?
(Correct answer: Program A; Intuitive answer: Program B).

PD3. No conflict version: The current recession has forced companies to cut jobs. Your office provides financial support to struggling businesses in the local economy, but limited resources force you to choose which businesses to assist. Program A saves 52 of the 560 jobs that would have other-wise been lost at Factory A. Program B saves 54 of the 60 jobs that would have otherwise been lost at Factory B. These programs are mutually exclusive and the only two options available. Which would you choose?

PD4. An oil spill around Puget Sound is threatening the sea otter populations in two areas of the bay. Two cleanup plans are proposed, but there is only enough money to support one plan. So, there are only enough resources to save otters in one of these areas of the bay. Program A will save 124 of the 800 otters near the north end of the bay. Program B will save 120 of the 150 otters near the south end of the bay. These programs are mutually exclusive and the only two options available. Which would you choose?
(Correct answer: Program A; Intuitive answer: Program B).

PD5. A species of plant found only in a remote area of New Guinea is threatened with extinction by a recently introduced species of vine. You have access to two treatments that kill the vines and save the plants, but you only have enough money to fund one program. If you implement Program A, you will save 160 of the 200 plants located in Quadrant A. If you implement Program B, you will save 164 of the 820 plants located in Quadrant B. These programs are mutually exclusive and the only two options available.
(Correct answer: Program B; Intuitive answer: Program A)

Disjunctive Reasoning

DR1. Jack is looking at Ann but Ann is looking at George. Jack is married but George is not. Is a married person looking at an unmarried person? a) Yes, b) No, c) Cannot be determined. (Correct answer: a) Yes; Intuitive answer: c) Cannot be determined).

DR2. There are 5 blocks in a stack, where the second one from the top is green, and the fourth is not green. Is there a green block directly on top of a non-green block? a) Yes, b) No, c) Cannot be determined.
(Correct answer: a) Yes; Intuitive answer: c) Cannot be determined).

DR3. A lawyer sent an email to an engineer but the engineer sent an email to a doctor. The lawyer is American but the doctor is not. Did an American person send an email to a non-American person? a) Yes, b) No, c) Cannot be determined.
(Correct answer: a) Yes; Intuitive answer: c) Cannot be determined).

DR4. Sue called Sara and Sara called Jim. Sue is blonde but Jim is not. Did a blonde person call a non-blonde person? a) Yes, b) No, c) Cannot be determined.
(Correct answer: a) Yes; Intuitive answer: c) Cannot be determined).
DR4. No conflict version: Sue called Sara and Sara called Jim. Sara is blonde but Jim is not. Did a blonde person call a non-blonde person? a) Yes, b) No, c) Cannot be determined.
(Correct and intuitive answer: a) Yes).

DR5. There are 5 persons in a queue. The second person in the queue is a smoker and the fourth is a non-smoker. Is there smoker next to a non-smoker in the queue? a) Yes, b) No, c) Cannot be determined.
(Correct answer: a) Yes; Intuitive answer: c) Cannot be determined).

Cognitive Reflection Test (adapted version)

CRT1. James and Sue started a joint savings account by depositing 1200€. James deposited 1000€ more than Sue. How much did Sue deposit?
CRT2. If 15 bakers bake 15 cakes in 15 minutes, how long would it take 30 bakers to bake 30 cakes?

CRT3. A computer virus is spreading through the system of a computer. Every minute, the number of infected files doubles. If it takes 100 minutes for the virus to infect all of the system, how long would it take for the virus to infect half of the system?

C1.2. Illustration of other participant's profiles used in Experiment 1

High competence condition

Profile picture:

Name (optional): David

Age: 35

Gender: Male

Occupation: Artificial Intelligence and Robotics Consultant

Hobbies: Reading sci-fi books, building drones, programming, watching series, ...

Low competence condition

Profile picture:

Name (optional): Sarah

Age: 19

Gender: Female

Occupation: Babysitter

Hobbies: Shopping at the mall, partying with friends, makeup & fashion events, ...

C1.3. Instructions used in Experiment 1

C1.3.1. General Instructions

Welcome!

This is a study about how people answer reasoning problems in social contexts.

Next, you'll be paired with another participant with whom you will work in the first part of this study.

The study starts by randomly choosing the roles played by each participant in the first block of trials: either respondent or evaluator.

In each trial, the participant who plays the role of respondent will be asked to answer a reasoning problem. The participant who plays the role of evaluator will then be asked to evaluate the provided answer, stating if it is correct or incorrect, and to provide his/her own answer to the same problem.

Both participants should give their answers and evaluations based on their own knowledge, without searching the web for the correct answers. It is extremely important that participants in either role respect this set of instructions – this study depends on it.

####

In order to introduce you to the other participant, before starting the study, you'll be asked to:

a) provide some demographic information (age, gender, occupation)

b) select a profile picture from a pre-defined set of options that best matches your personality and intellect (For confidentiality reasons we do not use participants' real pictures)

The other participant will do the same, so that you can also have some background information about him/her.

Name (optional):

Age:

Gender:

Occupation (if you are a student, please specify the area):

Hobbies:

####

Please select a profile picture that best matches your personality and intellect:

####

Please answer the following problems while you'll being paired with another participant in this study. [Adapted version of the Cognitive Reflection Test (Frederick, 2005)]

1. James and Sue started a joint savings account by depositing 1200€. James deposited 1000€ more than Sue. How much did Sue deposit?
2. If 15 bakers bake 15 cakes in 15 minutes, how long would it take 30 bakers to bake 30 cakes?
3. A computer virus is spreading through the system of a computer. Every minute, the number of infected files doubles. If it takes 100 minutes for the virus to infect all of the system, how long would it take for the virus to infect half of the system?

C1.3.2. Instructions used in the first block of problems

First Block

You were assigned the role of EVALUATOR and David/[vs. Sarah] (the other participant in the study) will be the respondent.

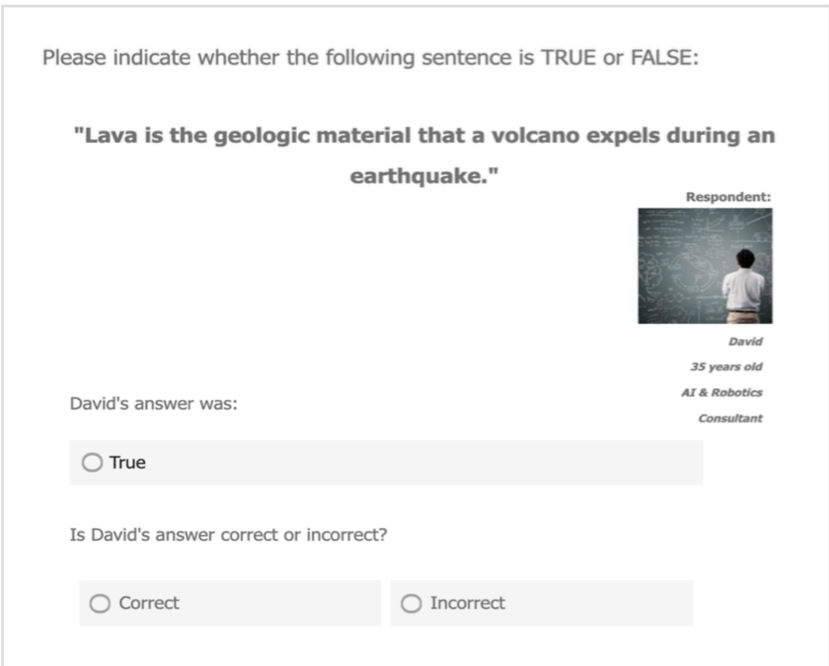
In the following trials, David/[vs. Sarah] will be asked to answer several reasoning problems.

Your task will be to evaluate David's/[vs. Sarah] answers, stating whether they are correct or incorrect, and to give your own answers to the same reasoning problems, stating how confident you are about your answers.

In each problem, you should:

- indicate whether David's/[Sarah's] answer is correct or incorrect. If you think that David's/[Sarah's] answer is incorrect, you should provide your own answer to the presented problem;
- state how confident you are about your answer on a scale ranging from 1 (not at all confident) to 9 (totally confident).

C1.3.3. Illustration of the trials


Please indicate whether the following sentence is TRUE or FALSE:

"Lava is the geologic material that a volcano expels during an earthquake."

Respondent:

David
35 years old
AI & Robotics
Consultant

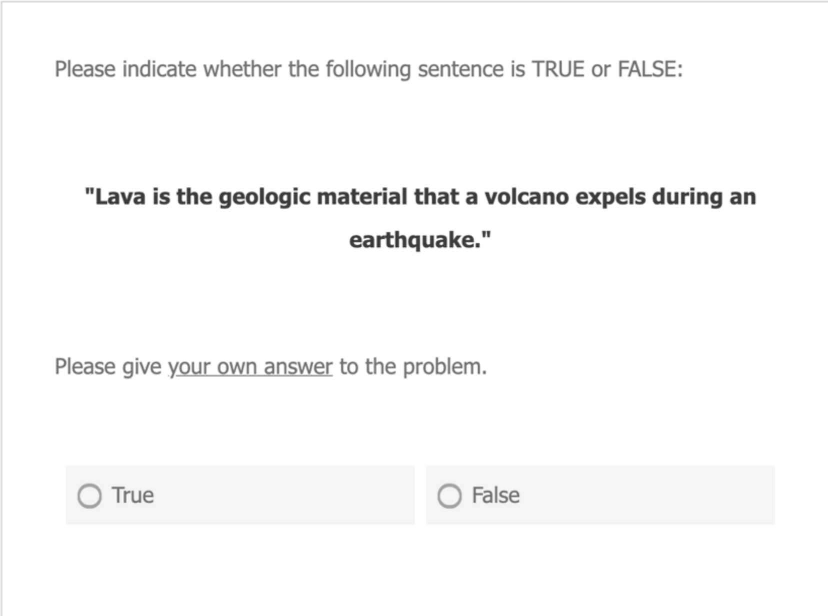
David's answer was:

◯ True

Is David's answer correct or incorrect?

◯ Correct ◯ Incorrect

###



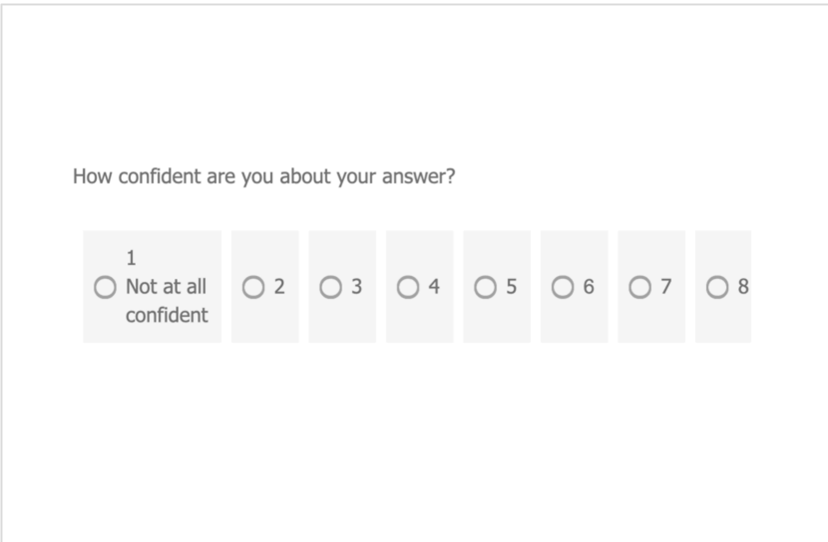
Please indicate whether the following sentence is TRUE or FALSE:

"Lava is the geologic material that a volcano expels during an earthquake."

Please give <u>your own answer</u> to the problem.

◯ True ◯ False

###



How confident are you about your answer?

◯ 1 Not at all confident ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8

<u>C1.3.4. Instructions used in the second block of problems</u>

Second block

You have finished the first part of the study!
Next, you'll be asked to answer another 5 reasoning problems, but this time you'll be working alone.

In each problem, you should:

- answer as accurately as possible;
- state how confident you are about your answer on a scale ranging from 1 (not at all confident) to 9 (totally confident).

C1.4. Evaluation of groups task

You have finished the second part of the study!

In the next part of the study, you'll be presented with several occupational groups together with different characteristics. For each occupational group you'll be asked to indicate to what extent the group has each of the characteristics. Please note that there are no correct answers! We are interested in your first impressions, so please give your answers without thinking too much.

Critical groups: AI & Robotics consultants, Babysitters
Fillers: University Professors, Doctors without borders, Fashion hairstylists, and Social Workers

How SMART are [GROUP]?
How RATIONAL are [GROUP]?
How ANALYTICAL are [GROUP]?
How KIND are [GROUP]?
How SENSITIVE are [GROUP]?
How WARM are [GROUP]?

Rating scale anchored on 1 – Not at all to 7 – Totally

C2.1. Reasoning problems used in Experiment 2

Syllogisms

SP1. Believable-invalid

Premissa 1: Todas as flores precisam de água.

Premissa 2: As rosas precisam de água.

Conclusão: Logo, as rosas são flores.

SP2. Believable-invalid

Premissa 1: Todos os animais têm olhos.

Premissa 2: Os gatos têm olhos.

Conclusão: Logo, os gatos são animais.

SP3. Unbelievable-valid

Premissa 1: Todos os mamíferos conseguem andar.

Premissa 2: As baleias são mamíferos.

Conclusão: Logo, as baleias conseguem andar.

SP4. Unbelievable-valid

Premissa 1: Todos os peixes voam.

Premissa 2: A sardinha é um peixe.

Conclusão: Logo, a sardinha voa.

SP5. Believable-invalid

Premissa 1: Todos os aviões têm asas.

Premissa 2: O Boeing 747 tem asas.

Conclusão: Logo, o Boeing 747 é um avião.

SP6. Unbelievable-invalid (no conflict version)

Premissa 1: Todos os pássaros têm asas .

Premissa 2: Os cães têm asas.

Conclusão: Logo, os cães são pássaros.

ST1. Believable-invalid

Premissa 1: Todas os veículos motorizados poluem o ambiente.

Premissa 2: Os skates não são veículos motorizados.

Conclusão: Logo, os skates não poluem o ambiente.

ST2. Believable-invalid

Premissa 1: Todos os cães são ágeis

Premissa 2: As tartarugas não são cães.

Conclusão: Logo as tartarugas não são ágeis.

ST3. Unbelievable-valid

Premissa 1: Todas as coisas com motor precisam de combustível.

Premissa 2: Os carros não precisam de combustível.

Conclusão: Logo, os carros não têm motor.

ST4. Unbelievable-valid

Premissa 1: Todas as bebidas têm álcool.

Premissa 2: A coca-cola não tem álcool.

Conclusão: Logo, a coca-cola não é uma bebida.

ST5. Unbelievable-Valid

Premissa 1: Todos os planetas são redondos

Premissa 2: A Terra não é redonda.

Conclusão: Logo a Terra não é um planeta.

ST6. Believable-valid (no conflict version)

Premissa 1: Todos os eletrodomésticos consomem energia.

Premissa 2: As cadeiras não consomem energia.

Conclusão: Logo, as cadeiras não são eletrodomésticos.

Transitive reasoning

RT1. Believable-invalid

Premissa 1: Os elefantes são maiores do que as bicicletas.

Premissa 2: Os zurros são maiores do que as bicicletas.

Premissa 3: Os ratos são maiores do que os zurros.

Conclusão: Logo, os elefantes são maiores do que os ratos.

RT2. Believable-invalid

Premissa 1: Os leões são mais ferozes do que os elefantes.

Premissa 2: Os rodes são mais ferozes do que os elefantes.

Pemissa 3: Os ratos são mais ferozes do que os rodes.

Conclusão: Logo, os leões são mais ferozes do que os ratos.

RT3. Unbelievable-valid

Premissa 1: Os candeeiros são mais luminosos do que os faróis.

Premissa 2: Os ripes são mais luminosos do que os faróis.

Premissa 3: As velas são mais luminosas do que os ripes.

Conclusão: Logo, as velas são mais luminosas do que os faróis.

RT4. Unbelievable-valid

Premissa 1: Os gatos são mais pesados do que os mudes.

Premissa 2: Os mudes são mais pesados do que os ursos.

Premissa 3: Os camelos são mais pesados do que os ursos.

Conclusão: Logo, os gatos são mais pesados do que os ursos.

RT5. Believable-invalid

Premissa 1: Os Renaults são mais baratos do que os Sunai.

Premissa 2: Os BMWs são mais baratos do que os Sunai.

Premissa 3: Os Mercedes são mais baratos do que os BMWs.

Conclusão: Logo, os Renaults são mais baratos do que os Mercedes.

RT6. Unbelievable-invalid (no conflict version)

Premissa 1: Os dias são mais longos do que os ernes.

Premissa 2: Os ernes são mais longos do que os segundos.

Premissa 3: Os minutos são mais longas do que os segundos.

Conclusão: Logo, os minutos são mais longos do que os dias.

RT7. Believable-invalid

Premissa 1: As baleias são maiores do que os potis.

Premissa 2: Os coelhos são menores do que os potis.

Premissa 3: As minhocas são maiores do que os coelhos.

Conclusão: Logo, as baleias são maiores do que as minhocas.

RT8. Believable-invalid

Premissa 1: Os elefantes são mais pesados do que os hipopótamos.

Premissa 2: Os mobes são mais leves do que os elefantes.

Premissa 3: Os mobes são mais pesados do que os hamsters.

Conclusão: Logo, os hipopótamos são mais pesados do que os hamsters.

RT9. Unbelievable-valid

Premissa 1: As revistas têm mais páginas do que os folhetos.

Premissa 2: Os trones têm menos páginas do que as revistas.

Premissa 3: Os trones têm mais páginas do que os romances.

Conclusão: Logo, as revistas têm mais páginas do que os romances.

RT10. Unbelievable-valid

Premissa 1: A Islândia é mais quente do que a Sérvia.

Premissa 2: A Pruslândia é mais fria do que a Islândia.

Premissa 3: A Pruslândia é mais quente do que o Brasil.

Conclusão: Logo, a Islândia é mais quente do que o Brasil.

RT11.Believable-valid (no conflict version)

Premissa 1: As lagartas são mais rápidas do que as térmitas.

Premissa 2: As lagartas são mais lentas do que os cavalos.

Premissa 3: Os cavalos são mais rápidos do que os toncos.

Conclusão: Logo, os cavalos são mais rápidos do que as térmitas.

RT12_ Unbelievable-valid

Premissa 1: As cidades são maiores do que as casas.

Premissa 2: Os toncos são menores do que as cidades.

Premissa 3: Os toncos são maiores do que os continentes.

Conclusão: Logo, as cidades são maiores do que os continentes.

<u>Cognitive Reflection Test (adapted version)</u>

Para terminar, pedimos-lhe que responda a um último bloco de problemas de raciocínio. Por favor, responda a todas as questões de forma tão correta quanto possível.

CRT1. O João e a Maria iniciaram uma conta poupança conjunta com 1200€. O João depositou 1000€ mais do que a Maria. Quanto dinheiro depositou a Maria?

CRT2. Quinze pasteleiros fazem 15 bolos em 15 minutos. Quanto tempo demorariam 30 pasteleiros a fazer 30 bolos?

CRT3. Um terreno está infetado com uma praga de insetos. A cada hora, a área de terreno infetada por esta praga duplica. Sabendo que a praga de insetos demora 50 horas para infetar todo o terreno, quanto tempo demoraria para infetar metade do terreno?

CRT4. Um tijolo pesa um quilo mais meio tijolo. Quanto pesa um tijolo?

CRT5. Está a participar numa corrida e ultrapassou a pessoa que estava em segundo lugar. Em que lugar ficou?

CRT6. Dois jogadores A e B têm 7 moedas cada um. Se o jogador A perder uma moeda para o jogador B, com quantas moedas a mais fica o jogador B?

C2.2. Instructions used in Experiment 2

C2.2.1. General instructions

Bem-vindo/a!

Neste estudo estamos interessados em saber mais sobre o modo como as pessoas resolvem problemas de raciocínio em contexto social.
Para tal, vamos apresentar-lhe de seguida um conjunto de problemas de raciocínio lógico. Para simular o contexto social, em cada problema vamos apresentar-lhe a resposta de outra pessoa que já respondeu anteriormente a este estudo. Estas respostas tanto podem estar certas como erradas.
A sua tarefa em cada ensaio será indicar a sua própria resposta ao problema apresentado.

Antes de começar o estudo, pedimos-lhe que nos indique, por favor:

Idade:
Sexo:

###

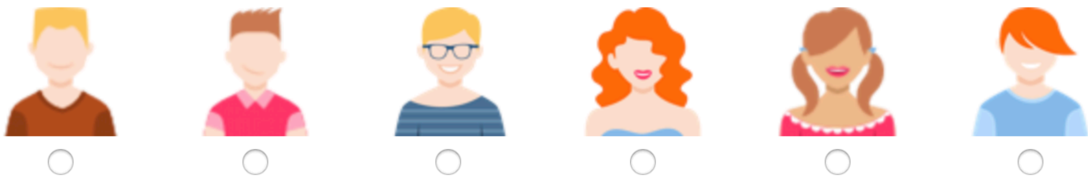
Autoriza que as suas respostas sejam apresentadas a outras pessoas que futuramente participem neste estudo?

Sim ___ Não ___

###

Caso autorize a partilha das suas respostas em estudos futuros, pedimos-lhe que preencha os campos abaixo com alguma informação que permita criar o perfil de participante que acompanhará as suas respostas.

Selecione um ícone para acompanhar as suas respostas:



Profissão (caso seja estudante indique, por favor, o seu curso):

Hobbies (sempre que possível especifique as suas preferências – por exemplo: cinema – filmes de ação):

C2.2.2. Instructions used in the first block of problems

Início do estudo

Vamos agora dar início ao estudo. Iremos apresentar-lhe 24 problemas de raciocínio lógico, acompanhados da resposta que o **Participante A** [Participante B] deu a esses problemas. Todos os problemas apresentam algumas premissas e uma conclusão. Em cada problema, a conclusão é válida se decorrer logicamente das premissas, caso contrário é logicamente inválida.
Por exemplo, considere o seguinte problema de raciocínio lógico:

Premissa 1: Todos os adultos correm.
Premissa 2: O João é adulto.
Conclusão: Logo, o João corre.

Se assumirmos que todos os adultos correm e que o João é adulto então podemos concluir que o João corre. Ou seja, a conclusão é válida porque decorre logicamente das premissas, independentemente da premissa 1 poder não corresponder ao conhecimento que temos do mundo.

A sua tarefa é avaliar, para cada problema, se a conclusão é logicamente válida ou não, e indicar o grau de confiança que tem na sua resposta.

C2.2.3. Illustration of the trials

Premissa 1: Todas as coisas com motor precisam de combustível.
Premissa 2: Os carros não precisam de combustível.
Conclusão: Logo, os carros não têm motor.

Assumindo que as premissas são verdadeiras, indique se a conclusão decorre logicamente das premissas.

25 anos
Análise Matemática
e Computação

A resposta do participante A foi:

Sim, a conclusão decorre logicamente das premissas.

Indique a sua resposta para o problema apresentado.

Sim, a conclusão decorre logicamente das premissas.

Não, a conclusão não decorre logicamente das premissas.

Quão confiante está na sua resposta?

1 Nada confiante	2	3	4	5	6	7 Totalmente confiante

C2.2.4. Instructions used in the second block of problems

Terminou a primeira fase do estudo. De seguida vamos apresentar-lhe um novo conjunto de problemas de raciocínio, mas desta vez não irá ver as respostas que o Participante A [Participante B] deu a esses problemas.

Assim, em cada ensaio a sua tarefa será:

- indicar a sua resposta ao problema apresentado;
- indicar quão confiante está na sua resposta.

C2.3. Illustration of other participant’s profiles used in Experiment 2

De seguida vamos apresentar-lhe as respostas de um dos seguintes participantes:

PARTICIPANTE A

Idade: 25 anos
Sexo: M

Profissão: Estudante de Doutoramento em Matemática Aplicada e Computação

Hobbies: Ler (ficção científica), jogos de lógica, programação, dar explicações de matemática

PARTICIPANTE B

Idade: 20 anos
Sexo: M

Profissão: Estudante de artes plásticas e design

Hobbies: beber copos com os amigos, DJ (electro-trance), fazer desenhos para tatuagens, ver séries de TV e viajar à boleia.

Com base nos respetivos perfis, estime por favor qual o desempenho de cada um dos participantes na resolução dos problemas de raciocínio lógico. Indique a sua estimativa usando a escala abaixo (1 – Muito mau a 7 – Muito bom).

Participante A: ______

Participante B: ______

Clique no botão com a seta para ficar a conhecer qual dos participantes (A ou B) lhe foi atribuído.